

```

FT MOD RES 1057 1057 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CONFLICT 25 25 P -> T (IN REF. 1).
FT CONFLICT 679 679 G -> D (IN REF. 3).
FT CONFLICT 783 784 LV -> VL (IN REF. 1).
FT CONFLICT 917 917 S -> C (IN REF. 1).
FT CONFLICT 1341 1367 ULTSCINGPWPAPPPTGHERGNA -> RSPPV
FT SEQUENCE 1367 AA; 152516 MM; EFC99704F1DCA266 CRC64;
SQ SEQUENCE 1367 AA; 152516 MM; EFC99704F1DCA266 CRC64;

Query Match 4.7%; Score 109; DB 1; Length 1367;
Best Local Similarity 21.0%; Pred. No. 12;
Matches 45; Conservative 36; Mismatches 73; Indels 60; Gaps 9;

QY 9 HLLVLLVLLLPATGKNVVLGKGDVLELTCTASOKSIQPHMKNNGIILNGSGF 68
DB 544 HIRGEIRIVQPAHQTEQ-----ESVSLCTARNTFENITWK-----LSQATS 590
QY 69 LTKGPKLNDRADSRSLW-----DOGNPLLI-KNLKIEDSTYICEVDQKEE- 117
DB 591 VHMGS-LTPVCNLDLWKNGTMSNSTNDILYAFGNASLDQGDYVCSADPKTK 649
QY 118 -----VQLVFGITANSDFHLQGSGLTLTLESPPSSSVQCRSPKX----- 162
DB 650 RHCLVQLIILERMAMITGNLENQTTT-----GETLEVTC--PASGNPTHTWFXD 701
QY 163 -----IQGKTLVSQLELDQSGTWC 184
DB 702 NETVEDSGIVLRDGNRLITRKVKEDGGLYTC 735

RESULT 72
PVR2 HUMAN STANDARD; PRT; 538 AA.
ID Q92692; O75455; Q96029;
AC 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DR 10-OCT-2003 (Rel. 42, Last annotation update)
DE Poliovirus receptor related protein 2 precursor (Herpes virus entry
GN mediator B) (HvEB) (Nectin 2) (CD112 antigen).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RX MEDLINE FROM N.A. (ISOFORM DELTA).
RA MEDLINE:95347610; PubMed:7622062;
RT "The human PRR2 gene, related to the human poliovirus receptor gene
RL Gene 159:267-272(1995).
RN [2]
RX MEDLINE FROM N.A. (ISOFORM ALPHA).
RA MEDLINE:98321161; PubMed:9657005;
RT Warner M.S., Geraghty R.J., Martinez W.M., Montgomery R.I.,
RA Whitbeck J.C., Xu R., Eisenberg R.J., Cohen G.H., Spear P.G.,
RT "A cell surface protein with herpesvirus entry activity (HvEB) confers
RT susceptibility to infection by mutants of herpes simplex virus type
RT 1, herpes simplex virus type 2, and pseudorabies virus.";
RL Virology 246:179-189(1998).
RN [3]
RX MEDLINE:22388257; PubMed:12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo T.L., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin B.B., Toshiyuki S., Carninci P., Pange C.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

```

```

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RX SEQUENCE OF 31-538 FROM N.A.
RA Yoshiura K., Murray J.C.;
RT "A transcriptional map in the region of 19q13 derived using direct
RT sequencing and exon trapping.";
RN Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RX SEQUENCE OF 449-538 FROM N.A.
RA MEDLINE:99449047; PubMed:10520737;
RA Fretz E.M., Zhang W.J., Lalonde J.P., Tay G.K., Gaudieri S.,
RA Ashworth L.K., Van Bockxmeer F.M., Dawkins R.L.;
RT "Sequencing of 42kb of the APO E-C2 gene cluster reveals a new gene:
RT PEREC1.";
RL DNA Seq. 9:89-101(1998).
CC -1- FUNCTION: RECEPTOR FOR ALPHAHHERPESVIRUS (HSV-1, HSV-2 AND
CC PSEUDORABIES VIRUS) ENTRY INTO CELLS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Delta;
CC IsoId=Q92692-1; Sequence=Displayed;
CC Name=Alpha;
CC IsoId=Q92692-2; Sequence=VSP_002628, VSP_002629;
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.
CC -1- DATABASE: NAME=PROV; NOTE=PROV 1:74-77(2000);
CC WWW="http://www.ncbi.nlm.nih.gov/prov/guide/204270028_g.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X80038; CAA56342.1; -
DR EMBL; AF058448; AAC23797.1; -
DR EMBL; BC003091; AAH03091.1; -
DR EMBL; AF044968; AAC82348.1; -
DR EMBL; AF044962; AAC82348.1; JOINED.
DR EMBL; AF044963; AAC82348.1; JOINED.
DR EMBL; AF044964; AAC82348.1; JOINED.
DR EMBL; AF044965; AAC82348.1; JOINED.
DR EMBL; AF044967; AAC82348.1; JOINED.
DR EMBL; AF050154; AAD02503.1; -
DR PIR; I68093; I68093.
DR Genew; HGNC:9707; PVR2.
DR MIM; 600798; -
DR GO; GO:0005886; C:plasma membrane; TAS.
DR GO; GO:0015026; F:coreceptor activity; TAS.
DR InterPro; IPR007110; IG-1-like.
DR InterPro; IPR003599; IG.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00409; IG; 1.
DR PROSITE; PS50835; IG_LIKE; 3.
KW Immunoglobulin domain; Receptor; Transmembrane; Glycoprotein; Signal;
KW Repeat; Alternative splicing.
KW SIGNAL.
FT CHAIN 1 31 538 POTENTIAL.
FT CHAIN 32 538 POLIOVIRUS RECEPTOR RELATED PROTEIN 2.
FT DOMAIN 32 360 EXTRACELLULAR (POTENTIAL).

```

FT	TRANSMEM	361	381	POTENTIAL.
FT	DOMAIN	362	538	CYTOSOLSMIC. (POTENTIAL).
FT	DOMAIN	32	156	IG-LIKE V-TYPE.
FT	DOMAIN	162	256	IG-LIKE C2-TYPE 1.
FT	DOMAIN	261	345	IG-LIKE C2-TYPE 2.
FT	DISULFID	54	140	BY SIMILARITY.
FT	DISULFID	183	238	BY SIMILARITY.
FT	DISULFID	283	329	BY SIMILARITY.
FT	CARBOHYD	137	137	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	334	324	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	351	479	PNAGAGATGIIIGLIIAIIATAVAATGILLICQCRKEQTLL OAEATEDLEGGSYPKPPAKLEAOEMPSOLFUGASEH SPLKTPYFDGASCTEOEMPRIHELPTLBERSGPLRGATS LGSPRT -- RASPRVGVPLWGAAGVTGLIVLLIANGSLA FTLLRVRRRRSPPGAGGASDGGFYDKRAQVLGGSDVP WTFDPVGPMEPDGCKEEBEEBEBAEKAGMLPEPPLBEDM ESOLDLSIRRAAYV (in Isoform Alpha). FTTD-vsp 002628. /FTId-vsp 002629.
FT	VARSPLIC	480	538	Miasing (In isoform Alpha).
FT	SEQUENCE	538 AA,	57742 MW;	JAE4F83E92F6Fe24 CRC64;
SQ	SEQUENCE	538 AA,	57742 MW;	JAE4F83E92F6Fe24 CRC64;
<hr/>				
Query Match				
Beat Local Similarity 4.6%; Score 108; DB 1; Length 538;				
Matches 112; Conservative 21.3%; Pred. No. 4.4; Mismatches 204; Indels 152; Gaps				
Oy	2	NRGVFRHL	LVALLPAA	TQGNKV-----VLGRKDVTIELTC----- 41
Dd	10	SRSPPTPLLM	PIRLLLLLLETGA	QDVAVQLPEVRGDLGTVELPCHLPPVGLYISLV 69
Oy	42	-----TASOKS	IQFMKNSNOI---	KLNGSGSLTYTGPKSLNDRASRSIMWGKF 92
Dd	70	WORPADAPNHQ	VAAFHPPMGPS	FPSFKGSSRLSEFS--AKOSTGOOTELELODA--- 123
Oy	93	PLIKNLKI	EDDFYICEVED--	OKEVEQ---LVVGLTRANSPDTHLQGCSITLTLESPP 147
Dd	124	TLAHGLIVED	GNTCEPATRPKS	VGRCMTLRVIAKRNV-----QAEAKVTTSODP 177
Oy	148	GSSPSVQC	RSPRGK--NIQ	GKTLSVSQLELDSGSTWTCTYLONOKVEFKIDIIVLAF 204
Dd	178	--TTVALCISK	EGRPARISMWS	LIDWEAKETQVSGTLAGTV-----TVTSRRTLVPQSR 230
Oy	205	QKASSIVYKK	EEQVE--FSFLA	FVF-----KLTGSELWMQAE-----RASS 248
Dd	231	ADGVTVTCK	HESEFEERPALIV	TVTSRVPREVISISGYDMNYLGRDTATLSCDVASNPE 290
Oy	249	KSWITFDL	KNKVESVKRV	TQDPKIQMKYLPY-----LTLPAQL----- 288
Dd	291	PTGYMWSTTS	GCFPFPSAVAQ----	GSDLVIHANDLSLNTTFVCTVTNAVGKRAEYOYI 344
Oy	289	-----POYAGS	GNLTALAEAKTG	KHQEVNLVMRATOLOKNLTC----- 328
Dd	345	FVRETPNTA	GAG-----ATG	GIIGGLIAIIATVAATGILLICQCRKEQTLOGAEED 397
Oy	329	-EWGPTSPK	MLSLKLENKA	KVSKREKPVWLNDPAGMQCLLS--DSGCVLIESNI 384
Dd	398	EDLEGRP-----	SYKPPPKAKL	EAQEMPSOLFUGASEHSPLKTPYFDGASCTEDEM 451
Oy	385	---KVLPTW---	STPVHPASAL	APRNCSALPDQETASLRDDPA 424
Dd	452	PRYHELPTL	BERSGRPHCATSL-	---GSPFVFP-----DEPPA 486

OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 CC NCBI\_TaxId=9606;  
 CC [1]  
 RN RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RX MEDLINE=21303047; PubMed=11409878;  
 RA Yousef G.M., Oordon M.H., Fousheas G., Diamandis E.P.;  
 RT "Molecular characterization, tissue expression, and mapping of a novel  
 RT Sigslec-like gene (Slc2) with three splice variants."  
 RL Biochem. Biophys. Res. Commun. 284:900-910(2001).  
 RN [2]  
 RN RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Spleen;  
 RX MEDLINE=21181594; PubMed=11284738;  
 RA Munday J., Kerr S., Ni J., Cornish A.L., Zhang J.Q., Nicoll G.,  
 RA Flunday H., Mattei M.-G., Moore P., Liu D., Crocker P.R.;  
 RT "Identification, characterization and leucocyte expression of  
 RT Sigslec-10, a novel human sialic acid-binding receptor."  
 RL Biochem. J. 355:489-497(2001).  
 RN [3]  
 RN RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Dendritic cell;  
 RX MEDLINE=21359381; PubMed=11358961;  
 RA Li N., Zhang W., Wan T., Zhang J., Chen T., Yu Y., Wang J., Cao X.;  
 RA "Cloning and characterization of Sigslec-10, a novel sialic acid"  
 RT "binding member of the Ig superfamily, from human dendritic cells";  
 RL J. Biol. Chem. 276:28106-28112(2001).  
 RN [4]  
 RN RP SEQUENCE FROM N.A. (ISOFORM 3). ALTERNATIVE SPLICING (ISOFORMS 4  
 AND 5), MUTAGENESIS OF TYR-667, AND INTERACTION WITH PTPN6.  
 RX MEDLINE=22152890; PubMed=12163025;  
 RA Kitzig F., Martinez-Barriocanal A., Lopez-Botet M., Sayos J.;  
 RT "Cloning of two new splice variants of Sigslec-10 and mapping of the  
 RT interaction between Sigslec-10 and SHP-1."  
 RL Biochem. Biophys. Res. Commun. 296:355-362(2002).  
 RN [5]  
 RN RP SEQUENCE OF 334-697 FROM N.A. (ISOFORM 1/3/4).  
 RP TISSUE=B-cell;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins P.S., Wagner L., Sherman C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buettow K.H., Scheeter C.F., Hat F.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Sleptchenko M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Bernstein M.J., Ustin T.B., Toshynski S., Carninci P., Prange C.,  
 RA Baha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.O., McKernan K.O., Malek J.A., Gunaratne P.H.,  
 RA Richards S.K., Worley K.C., Hale S.S., Garcia A.M., Gay L.J., Huilyk S.W.,  
 RA Villalón D.K., Morley D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Ketterman M., Madan A.C., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shvachenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Buttefield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Scheraga A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC [1]- FUNCTION: Putative adhesion molecule that mediates sialic-acid  
 CC dependent binding to cells. Preferentially binds to alpha2,3- or  
 CC 2,6-linked sialic acid. The sialic acid recognition site may be  
 CC masked by cis interactions with sialic acids on the same cell  
 CC surface. In the immune response, may act as an inhibitory receptor  
 CC upon ligand induced tyrosine phosphorylation by recruiting  
 CC cytoplasmic phosphatase(s) via their SH2 domain(s) that block  
 CC signal transduction through dephosphorylation of signaling  
 CC molecules.  
 CC [1]- SUBUNIT: Interacts with PTPN6/SHP-1 upon phosphorylation.  
 CC [1]- SUBCELLULAR LOCATION: Type I membrane protein (isoforms 1-4);  
 CC secreted (isoform 5).  
 CC [1]- ALTERNATIVE PRODUCTS:

```

CC Event=Alternative splicing; Named isoforms=5;
CC Name=1; Synonyms=Long;
CC IsoId=Q96LC7-1; Sequence=Displayed;
CC Name=2; Synonyms=Short, SV1;
CC IsoId=Q96LC7-2; Sequence=VSP_002565;
CC Name=3; Synonyms=SV3;
CC IsoId=Q96LC7-3; Sequence=VSP_002564;
CC Name=4; Synonyms=SV4;
CC IsoId=Q96LC7-4; Sequence=VSP_002561;
CC Name=5; Synonyms=SV2;
CC IsoId=Q96LC7-5; Sequence=VSP_002562, VSP_002563;
CC TISSUE SPECIFICITY: Expressed by peripheral blood leukocytes
CC (eosinophils, monocytes and a natural killer cell subpopulation).
CC Isoform 5 is found to be the most abundant isoform. Found in lymph
CC node, lung, ovary and appendix. Isoform 1 is found at high levels
CC and isoform 2 at lower levels in bone marrow, spleen and spinal
CC chord. Isoform 2 is also found in brain. Isoform 4 is specifically
CC found in natural killer cells.
CC -1- DOMAIN: Contains 1 copy of a cytoplasmic motif that is referred
CC to as the immunoreceptor tyrosine-based inhibitor motif (ITIM).
CC This motif is involved in downmodulation of cellular responses.
CC The phosphorylated ITIM motif binds to the SH2 domain of
CC PTPN6/SHP-1.
CC -1- PTM: Tyrosine phosphorylated.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. SIGLEC
CC (sialic acid binding Ig-like lectin) family.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AY029277; AAK40255.1; -
DR EMBL; AY029277; AAK40256.1; -
DR EMBL; AF310233; AAK55139.1; -
DR EMBL; AF311905; AAK92542.1; -
DR EMBL; AY032685; AAK51124.1; -
DR EMBL; BC009955; AAK09955.1; -
DR Genew; HGNC:15620; SIGLEC10.
DR MIM; 606091; -
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003598; Ig_C2.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PRF00047; Ig_3.
DR SMART; SM00409; Ig_4.
DR SMART; SM00408; IgC2; 2.
DR PROSITE; PS00835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KW Cell adhesion; Lectin; Transmembrane; Signal; Glycoprotein;
KW Immunoglobulin domain; Repeat; Phosphorylation; Alternative splicing;
KW Polymorphism.
FT SIGNAL 1 16
FT CHAIN 1 697
FT DOMAIN 17 550
FT TRAMEM 551 571
FT DOMAIN 572 697
FT DOMAIN 18 121
FT DOMAIN 146 231
FT DOMAIN 251 339
FT DOMAIN 344 441
FT SITE 595 600
FT SITE 665 670
FT DISULFID 36 173
FT DISULFID 41 101
FT DISULFID 164 215
FT DISULFID 276 323
FT DISULFID 360 425

```

```

FT MOD_RES 667 667
FT FT
FT CARBOHYD 100 100
FT CARBOHYD 355 355
FT CARBOHYD 364 364
FT CARBOHYD 486 486
FT CARBOHYD 504 504
FT VARSPLIC 125 214
FT VARSPLIC 140 185
FT VARSPLIC 186 697
FT VARSPLIC 141 198
FT VARSPLIC 445 539
FT MUTAGEN 667 667
FT CONFLICT 28 28
FT CONFLICT 134 134
FT CONFLICT 226 226
FT CONFLICT 334 336
FT CONFLICT 344 344
FT CONFLICT 440 440
FT CONFLICT 587 587
FT CONFLICT 625 625
SQ SEQUENCE 697 AA; 76619 MW; 6CB231CE4941D1B CRC64;

Query Match 4.6%; Score 108; DB 1; Length 697;
Beet local Similarity 18.7%; Pred. No. 6.1;
Matches 98; Conservative 53; Mismatches 166; Indels 206; Gaps 22;

8 RHLVLQIALLPRA---ATQGNKTVL-GKKSGDYELTCTASQKKSIOFHMKNQKILG 63
DB DIVISISRNTPALPBPQGNVPLVLEAQKQFRLICADDSQRPATLSW----- 288
QY 64 NQGSFLTGPCKLNDRADSRSLMDQGNFPLIKLKIEDSDTYICEVEDO----- 114
DB 289 -----VLQNRVLSSHPWGPRLGLELGVNAGDSGRYTCRAENLGSQORALD 337
QY 115 -----KEEVOLVLF-----GLTNSDTHLQSGSLT---TLSPGSSPSVQGRS 157
DB 338 LSVQPPENLRVWVSQANRTVLENLNGTSLPVLGGSLCLVCVTHSSPPA----- 388
QY 158 PRGNIOGKTLVSQO-----LELDQSGTWTCTVLQNKVEFKIDIVLAFQK 206
DB 389 -RLSMTQKGVLSFSQSPDPGVLELPRVQVHBEFTHAR----- 428
QY 207 ASSIVYKKEGEQVFPFLAFTVEKLTGSGELW-----WQAEBASSKSWITFDL 256
DB 429 -----HPLGSHVSLSLSVHS-PKLLGPSCSMAEBGLHCSCSQASPAFLRWML----- 478
QY 257 KKEYSVYKRVQDPEKLGKGLPLHLTPQLPQYAGS-GULTLALAKTKLQHEVNLV 315
DB 479 -GEELLEGNSQDSF-----EYTPSAGPWANSSLSLH---GGLSSGL--- 517
QY 316 VMRAIQLQKNLTCEVWGPSTPKMLSLKEKKAQVS----- 352
DB 518 -----RLRCANVNHGQSGSILOLPKKGILSTAFSNGAFLGIGTALLFLCLAL 568
QY 353 -----KREKRVVNLPEAGMOCCLSDSGVLLSNIKYLPT-----WSPVPH 395
DB 569 IIMKILPKRRTQTEPRPRFSRHSITL-----DYINVPYAGPLAQKRNQKATPNS 619
QY 396 PRASALPAPPGSALPDQGTAS-----ALDPPAASALP 429
DB 620 PRT--PLVP--GAPSPESKKNQKQYQLPSPPEKSTQAP 656

```

RESULT 74

ID	TITLE	HUMAN	STANDARD	PRT	1138 AA.
AC	P35530				
DT	01-JUN-1994	(Rel. 29, Created)			
DT	01-JUN-1994	(Rel. 29, Last sequence update)			
DT	28-FEB-2003	(Rel. 41, Last annotation update)			
DE	Tyrosine-protein kinase receptor Tie-1 precursor (EC 2.7.1.112).				
GN	TIE OR TIE1.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_Taxid:9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE:92195316; PubMed:1112667;				
RA	Pattanen J., Armstrong E., Maekelae T.P., Korhonen J., Sandberg M.,				
RA	Rehkonen R., Knuutila S., Huebner K., Aitalo K.;				
RT	"A novel endothelial cell surface receptor tyrosine kinase with				
RL	extracellular epidermal growth factor homology domains.";				
RN	Mol. Cell. Biol. 12:1698-1707(1992).				
RP	[2]				
REVISIONS:					
RA	Revisions J.M.;				
RL	Submitted (UTU-1993) to the EMBL/GenBank/DBJ databases.				
CC	-1- FUNCTION: Probable protein tyrosine-kinase transmembrane receptor.				
CC	-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein				
CC	phosphate.				
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.				
CC	-1- TISSUE SPECIFICITY: Specifically expressed in developing vascular				
CC	endothelial cells.				
CC	-1- SIMILARITY: Belongs to the Tyr family of protein kinases. The				
CC	subfamily.				
CC	-1- SIMILARITY: Contains 2 immunoglobulin-like C2-type domains.				
CC	-1- SIMILARITY: Contains 3 fibronectin type III domains.				
CC	-1- SIMILARITY: Contains 3 EGF-like domains.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
CC	EMBL; X60957; CAA43290.1; -.				
CC	HSSP; P11362; IEGF.				
DR	Genew; HGNC:11809; TIE.				
DR	MIM; 600222; -.				
DR	GO; GO:0005887; C:integral to plasma membrane; TAS.				
DR	GO; GO:0004714; P:transmembrane receptor protein tyrosine kin. . .; TAS.				
DR	GO; GO:0007428; P:mesoderm development; TAS.				
DR	GO; GO:0007165; P:signal transduction; TAS.				
DR	InterPro; IPR006209; EGF-like.				
DR	InterPro; IPR008957; FN_III-like.				
DR	InterPro; IPR003961; FN_III.				
DR	InterPro; IPR006210; IEGF.				
DR	InterPro; IPR007110; IG-like.				
DR	InterPro; IPR003599; IG.				
DR	InterPro; IPR002049; Laminin_EGF.				
DR	InterPro; IPR000719; Prot_kinase.				
DR	InterPro; IPR001245; Tyr_kinase.				
DR	InterPro; IPR008266; Tyr_kinase_AS.				
DR	Pfam; PF00008; EGF; 2.				
DR	Pfam; PF00041; fn3; 3.				
DR	Pfam; PF00047; Ig; 2.				
DR	Pfam; PF00069; kinase; 1.				
DR	PRINTS; PR00011; EGF_LAMININ.				
DR	PRINTS; PR00109; TYRKINASE.				
DR	ProDom; PD000001; Prot_kinase; 1.				
DR	SMART; SMO0181; EGF; 2.				
DR	SMART; SMO0060; FN3; 3.				
DR	SMART; SMO0409; IG; 1.				
DR	SMART; SMO0219; TYRK; 1.				

DR						PROSITE; PS00022; EGF_1; 3.
DR						PROSITE; PS0186; EGF_2; 3.
DR						PROSITE; PS50026; EGF_3; 2.
DR						PROSITE; PS50017; PROTEIN_KINASE_ATP_1.
DR						PROSITE; PS50011; PROTEIN_KINASE_DOM_1.
DR						PROSITE; PS00109; PROTEIN_KINASE_TYR_1.
KW						Receptor; Tyrosine-protein kinase; Transferase; Signal; ATP-binding;
KW						Repeat; EGF-like domain; Transmembrane; Immunoglobulin domain;
KW						Glycoprotein; Phosphorylation; Multigene family.
FT	SIGNAL	1	24		POTENTIAL	
FT	CHAIN	25	1138		POTENTIAL	
FT	DOMAIN	25	759		EXTRACELLULAR (POTENTIAL).	
FT	TRANSEM	760	784		CYTOSOLIC (POTENTIAL).	
FT	DOMAIN	785	1138		CYTOSOLIC (POTENTIAL).	
FT	DOMAIN	43	105		IG-LIKE C2-TYPE 1.	
FT	DOMAIN	214	256		EGF-LIKE 1.	
FT	DOMAIN	258	303		EGF-LIKE 2.	
FT	DOMAIN	305	345		EGF-LIKE 3.	
FT	DOMAIN	372	426		IG-LIKE C2-TYPE 2.	
FT	DOMAIN	446	540		FIBRONECTIN TYPE-III 1.	
FT	DOMAIN	543	639		FIBRONECTIN TYPE-III 2.	
FT	DOMAIN	643	744		FIBRONECTIN TYPE-III 3.	
FT	DOMAIN	839	1118		PROTEIN KINASE.	
FT	NP_BIND	845	853		ATP (BY SIMILARITY).	
FT	BINDING	870	870		ATP (BY SIMILARITY).	
FT	ACT_SITE	979	979		BY SIMILARITY.	
FT	CARBOHYD	83	83		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	161	161		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	503	503		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	596	596		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHD	709	709		N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	MOD_RES	1007	1007		PHOSPHORYLATION (AUTO-) (BY SIMILARITY).	
SO	SEQUENCE	1138 AA;	125089 MW;		3B42BE33678C58A1 CRC64;	

Query Match	4.6%;	Score 107.5;	DB 1;	Length 1138;
Best Local Similarity	20.9%;	Pred. No. 12;		
Matches	83;	Conservative 44;	Mismatches 136;	Indels 135; Gaps 18
QY	109	CEVEDOKEVOLLVEGLTANS DTHLLOGQSITLTLESBP-----GSSP-----SVQCRS	157	
DB	344	CEKSDRIIPQI-----LNNASELEFNLEMTPRINCAAGNPFVPGSIELRK	389	
QY	158	PRGNIOGGKTL-----SVSQLELDSDGWTCTVLQN--QKVEFKIDIVYLAFAQ	205	
DB	390	PDGTVLLSTKAIVEPEKTTAEFEVPRVLVLADSGFEGECVSTSGQDSRRFRVNVKVPPEVP	449	
QY	206	KASSIVKKEBEOYFSPPLAFTVEKXTLGSGLMOMQABRASSKSWIT-----F	254	
DB	450	LAAPRLTLTKSORQVLVSPVLSFSGDGPISITVRLHLYRP--DSTMDWSTIVVDPSSENTLM	507	
QY	255	DLKNN-----EVSYKR-----VTOD--PKLQM-----GKKLPLHLT	283	
DB	508	NLRKRTGVSVRVQSLRSGEGEGEAGMGPPTLMTTTCPEBLLQPMLEGMHVBEGTDRLRVMS	567	
QY	284	LPGQALPOYAGSGNLTLLAEKTKLGHQEVNLVNRATQLOKULT-----CEV	330	
DB	568	LPLVGPGLVGGFILLRLMDGTRGGERRENVSSPOARTALLTGLTPGTHYQLDVOLYHCTL	627	
QY	331	WGPSPSLMLSLKLENKEAKVSKREKPVWTLNPAG-----MMQCLSDSGCVILLESNI	384	
DB	628	LGPASP-----PAHYVLLPSPGAPRPHLHAQALSDS-----	658	
QY	385	KVLPTWSTPVHPRASALPAPPTGSALPDPOFASALPDP	422	
DB	659	EIQLTWK--HF--EALPG-PISKTYVEVQVAGAGADP	650	





RL Biochem. Cell Biol. 66:1089-1099(1988).  
 RP [3]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=AX2.  
 RX MEDLINE=92406924; PubMed=1326559;  
 RA Desbaresats L., Lam T.Y., Wong L.M., Siu C.H.;  
 RT "Identification of a unique cAMP-response element in the gene  
 RT encoding the cell adhesion molecule gp80 in Dictyostelium  
 RT discoidium.";  
 RL J. Biol. Chem. 267:19655-19664(1992).  
 RN [4]  
 RN SEQUENCE OF 20-49.  
 RA Wong L.M., Siu C.H.;  
 RT "Cloning of cDNA for the contact site A glycoprotein of Dictyostelium  
 RT discoidium.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:4248-4252(1986).  
 RP [5]  
 RN CELL-BINDING DOMAIN.  
 RX MEDLINE=89034443; PubMed=3182938;  
 RA Kamboj R.K., Wong L.M., Lam T.Y.,  
 RT "Mapping of a cell-binding domain in the cell adhesion molecule gp80  
 RT of Dictyostelium discoidium.";  
 RL J. Cell Biol. 107:1835-1843(1988).  
 RN [6]  
 RP GPI-ANCHOR.  
 RX MEDLINE=89251561; PubMed=2721485;  
 RA Stadler J., Keenan T.W., Bauer G., Gerisch G.;  
 RT "The contact site A glycoprotein of Dictyostelium discoidium carries  
 RT a phospholipid anchor of a novel type.";  
 RL EMBO J. 8:371-377(1989).  
 CC -1- FUNCTION: This cell-surface glycoprotein mediates cell-cell  
 CC binding via homophilic interaction.  
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
 CC that contains a phosphoceramide moiety. Such anchor mediates a  
 CC fast and long persistence cell adhesion of the protein.  
 CC -1- DEVELOPMENTAL STAGE: Restricted to the aggregation stage of  
 CC development of D. discoidium.  
 CC -1- PTM: Acetylated, phosphorylated on serine and N-glycosylated with  
 CC two types of oligosaccharide chains.  
 CC -1- MISCELLANEOUS: The expression of this stringently regulated  
 CC protein during cell development is mediated through cell-surface  
 CC cAMP receptors.  
 CC -1- SIMILARITY: THE C-TERMINAL REGION CONTAINS CLUSTERS OF PROLINE  
 CC REGULARLY ALTERNATING WITH A HYDROXYAMINO ACID AND SHOWS  
 CC SIMILARITY TO THE HINGE REGION OF IMMUNOGLOBULINS. THIS DOMAIN  
 CC MIGHT ACT AS A SPACER TO ELEVATE SITES ACTIVE IN CELL CONTACT  
 CC INTO THE EXTRACELLULAR SPACE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: X04004; CAA27634.1; -;  
 DR EMBL: M36545; AAA33212.1; -;  
 DR EMBL: X66483; CAA47110.1; -;  
 DR PIR: S22066; A31643.  
 DR DictyBase; DB0001860; csaA.  
 DR InterPro; IPR002909; IPT\_TIG.  
 DR Pfam; PF01833; TIG; 2.  
 DR Cell adhesion; Glycoprotein; Membrane; Phosphorylation; Repeat;  
 KM GPI-anchor; Lipoprotein; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 492 CONTACT SITE A PROTEIN.  
 FT PROPEP 493 514 REMOVED IN MATURE FORM (POTENTIAL).  
 FT DOMAIN 20 453 GLOBULAR (POTENTIAL).  
 FT DOMAIN 454 485 PRO-RICH (HINGE STRUCTURE).  
 FT DOMAIN 462 479 2 X 8 AA REPEATS, PRO-RICH.  
 FT REPEAT 462 469 1.  
 FT REPEAT 472 479 2.

FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 137 137 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT LIPID 492 492 GPI-like-anchor amidated serine  
 FT (potential).  
 FT CONFLICT 216 216 V -> G (IN REF. 1).  
 FT SEQUENCE 514 AA; 53696 MW; 96A1C7CF42FEC096 CRC64;  
 SQ  
 Query Match 4.6%; Score 107; DB 1; Length 514;  
 Best Local Similarity 22.4%; Pred. No. 4.8;  
 Matches 102; Conservative 52; Mismatches 164; Indels 138; Gaps 24;  
 QY 63 GNQSSFLTKGSPKLNDRADRSRLMDQGNPPLIKLNKIDS-DPIYCEVEDQKEEYQL 121  
 DB 116 GRIGMILVDGFSNISC-----YKLVNDSDINSAMLSV--TDSVSPT 155  
 QY 122 VFGILTANSDTHLQGGSLTLTLSPGSSPSVQCRS-----PRGRNIQ 164  
 DB 156 IYFLVPNT-----IAGLILNELIQPFQFSTIVTSKVSFPTITSITPLARDLFTVT 211  
 QY 165 GG---KTLVSQ-----LELDGCTWCTVQLONQKV-EFKIDIVYLAFOKASSIVYK 213  
 DB 212 GKFEVTTASVTMGSHIYGLTVDDGT-NCHVIFTRSVESNTITPAKASTGVDMYLD 270  
 QY 214 KEQGEVFEFSF-----PLAFTEKLTGSGELMWQBRASSSMWITFDLKNKEVSKVATOD 269  
 DB 271 NQGNQGPITTTYPNPPTTSTKQVNDSEI-----STWTGDF--TQISLTMTSS 319  
 QY 270 PK--LQMGKKLPLHLTPQALPO-----YAGSGN-----LTLLAEKTKGLHQ-- 310  
 DB 320 PTNLVITGTBEKIYITLPHALPEGEIQENLKAGISNVVTSLTAVPINSVQAHPHNGS 379  
 QY 311 -----EVNLYMRAITQLOKNTCEWQPTSKMLSLKEMKEAKVSKREKRW 359  
 DB 380 ITTSGIFLNNAHVSIIV-----DQNTTIDVCAQDS-----NGESITCEVEAGSG 423  
 QY 360 VLNPEAGMOCGLSDSGOVLLESNIKVLPTWSTPV-----HPRASALPAP-PTGSLALPD 412  
 DB 424 TINLVNTNKNFASD-----PIKTEATSTTITTPDPTPTDVTATPSPTPIETATPS 476  
 QY 413 P-----QTASALPDPPASAL--PALAV--ISFL 439  
 DB 477 PTPKPTSTPTEPTEPSSATLISPLSIVIFISVL 512  
 RESULT 77  
 CEAL MOUSE STANDARD; PRT; 521 AA.  
 ID CEAL MOUSE  
 AC P31809;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Carcinoembryonic antigen-related cell adhesion molecule 1 precursor  
 DE (Biliary glycoprotein 1) (BGP-1) (Murine hepatitis virus receptor)  
 DE (HAV-R) (Biliary glycoprotein D).  
 GN CEACAM1 OR BGP OR BGPI OR BGPD.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CD-1; TISSUE=Colon.  
 RX MEDLINE=93100785; PubMed=8380065;  
 RA Dveksler G.S., Dieffenbach C.B., Cardelliichio C.B., McCaig K.,  
 RT Pensiero M.N.;  
 RT "Several members of the mouse carcinoembryonic antigen-related  
 RT glycoprotein family are functional receptors for the coronavirus  
 RT mouse hepatitis virus-A59.";  
 RL J. Virol. 67:1-8(1993).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Liver;  
 RX MEDLINE=92046352; PubMed=1719235;  
 RA Drexler G.S., Pensiero M.N., Cardelli C.B., Williams R.K.,  
 RT Jiang G.-S., Holmes K.V., Dieffenbach C.W.;  
 "Cloning of the mouse hepatitis virus (MHV) receptor: expression in  
 human and hamster cell lines confers susceptibility to MHV.";  
 RT J. Virol. 65:6881-6891 (1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93273228; PubMed=8500759;  
 RT McCallig K., Rosenberg M., Nedellec P., Turbide C., Beauchemin N.;  
 "Expression of the Bgp gene and characterization of mouse colon  
 RT biliary glycoprotein isoforms";  
 RT Gene 127:173-183 (1993).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
 RC STRAIN=CD-1; TISSUE=Colon;  
 RX MEDLINE=89195121; PubMed=2702644;  
 RA Beauchemin N., Turbide C., Afar D., Raymond M., Bell J.,  
 RT Stanners C.P., Fuks A.;  
 "A mouse analogue of the human carcinoembryonic antigen.";  
 RT Cancer Res. 49:2017-2021 (1989).  
 RN [5]  
 RP SEQUENCE OF 35-59.  
 RX MEDLINE=91268498; PubMed=1648219;  
 RA Williams R.K., Jiang G.-S., Holmes K.V.;  
 RT "Receptor for mouse hepatitis virus is a member of the  
 RT carcinoembryonic antigen family of glycoproteins";  
 RT Proc. Natl. Acad. Sci. U.S.A. 88:5533-5536 (1991).  
 CC -1- FUNCTION: Unknown; receptor for murine coronavirus MHV-A59.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=2;  
 CC Name=Long;  
 CC IsoId=P31809-1; Sequence=Displayed;  
 CC Name=Short;  
 CC IsoId=P31809-2; Sequence=VSP\_002484, VSP\_002485;  
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily, CEA family.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL, X67279; CAA47696.1; -;  
 DR EMBL, M7196; AAA37858.1; -;  
 DR EMBL, X15351; CAA33409.1; -;  
 DR PIR, JCI505; WMSR1.  
 DR PIR, JCI508; JCI508.  
 DR MGI:1347245; Ceacam1.  
 DR InterPro: IPR007110; Ig-like.  
 DR Pfam: PF00047; Ig; 3.  
 DR PROSITE, PS50835; IG\_LIKE; 3.  
 KW Immunoglobulin domain; Glycoprotein; Transmembrane; Signal;  
 KW Repeat; Alternative splicing; Receptor.  
 FT SIGNAL 1 34  
 FT CHAIN 35 521  
 FT DOMAIN 35 428  
 FT TRAMEM 429 447  
 FT DOMAIN 448 521  
 FT DOMAIN 35 142  
 FT DOMAIN 147 234  
 FT DOMAIN 239 319  
 FT DOMAIN 323 411  
 FT DISULFID 167 217  
 FT DISULFID 261 301  
 PROBABLE.

FT DISULFID 346 394 PROBABLE.  
 FT CARBOHYD 71 71 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 89 89 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 199 199 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 226 226 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 294 294 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 304 304 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 455 458 GSDQ -> SGSE (in isoform Short).  
 FT VARSPLIC 459 521 Missing (in isoform Short).  
 FT VARSPLIC 521 521 Missing (in isoform Short).  
 SQ SEQUENCE 521 AA; 57015 MW; 1C8F71FAC47DD54E CRC64;  
 Query Match 4.6%; Score 107; DB 1; Length 521;  
 Best Local Similarity 18.4%; Pred. No. 4.9;  
 Matches 80; Conservative 63; Mismatches 142; Indels 150; Gaps 18;  
 QY 16 LALLPAATQGNKVKVIGKKGVDELCTA-SQKSIQFHWKNSQIKILOGSFLTKGPS 74  
 DB 143 ILKKNINISNNNSNPV-BGDDSVSLTCDSTYDPDININILMSRN-----GESLSEGD- 191  
 QY 75 KLNDADARRSLMOGNGFPLIINKIEDSDTYICEVE-----DQKEVQI-LVFG----- 124  
 DB 192 -----RLKLSGNRTLTLLVNRDTPGVCETRNPNVSVNSDPISLNIYGPPTP 242  
 QY 125 LTANSDFHLLOGSILTLTLSPSSPSSPVQC-----RSPRKNIQGKTLSSVSOLELDS 179  
 DB 243 IISPDILVHGNSNINLSCHA--ASNPPAQFVWLINERPHSS-----GELFIPNITNNS 296  
 QY 180 GTWTTTVLQNKQKVEFKIDIVLVAFQKASIVYKKEGQVEFSPPLAFTVEKLTGSGELW 239  
 DB 297 GYVTC-----FVNSSVTG----- 309  
 QY 240 WQAEASASSKMITFDLKNKEVSRYVQDPKLOMGKLLPLHLTPQALPQVAGSGLTL 299  
 DB 310 -----LSKRTV-----KNITVLEPYTQPLQV----- 331  
 QY 300 ALEAKTGKLDQEVMLVVRATQLOKNTLCEV-WGPTSPKMLSLKENKAKVSKREKPV 358  
 DB 332 -----TNTTVLELDSVTL--TCLSNIDIGANTQMLENSGLQTERMTLSQNNISIKRIDP- 383  
 QY 359 WVLNPEAGMOCCLSDSGOVLLESNIKVLPTWSPVPHRASALPAAPGSAAPPQRTASA 418  
 DB 384 -IKREDAGEYQCEISNPVSVRSNSIKL-----DIIFDPFGGLSDGAIAGI 429  
 QY 419 LPDPPASALPAAAL 433  
 DB 430 VIGVAVGVALIAGLA 444  
 RESULT 78  
 SCA4\_RICPR STANDARD; PRT; 1022 AA.  
 ID SCA4\_RICPR Q9ZD49; Q9ZD48;  
 AC Q9ZD49; Q9ZD48;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Antigenic heat-stable 120 kDa protein (P5120) (120 kDa antigen)  
 DE (Protein P5 120).  
 GN SCA4 OR RP498/RP499.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;  
 CC Rickettsiaceae; Rickettsiase; Rickettsia;





CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=93115512; PubMed=8392071;  
 RA Hendrix P., Mayer-Jaekel R.E., Cron P., Goris J., Hofsteenge J.,  
 RT Merlevede W., Hemmings B.A.;  
 "Structure and expression of a 72-kDa regulatory subunit of protein  
 phosphatase 2A. Evidence for different size forms produced by  
 alternative splicing."  
 RL J. Biol. Chem. 268:15267-15276(1993).  
 RN [2]  
 RP SEQUENCE OF 768-1150 FROM N.A.  
 RA Pluvinct R., Estivill X., Escarceller M., Smey L.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: The B regulatory subunit might modulate substrate  
 selectivity and catalytic activity, and also might direct the  
 localization of the catalytic enzyme to a particular subcellular  
 compartment.  
 CC -1- SUBUNIT: PP2A consists of a common heterodimeric core enzyme,  
 composed of a 36 kDa catalytic subunit (subunit C) and a 65 kDa  
 constant regulatory subunit (PR65 or subunit A), that associates  
 with a variety of regulatory subunits. Proteins that associate  
 with the core dimer include three families of regulatory subunits  
 B (the R2/B/PR55/B55, R3/B'/PR72/PR130/PR59 and R5/B''/556  
 families), the 48 kDa variable regulatory subunit, viral proteins,  
 and cell signaling molecules.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=PR130; Synonyms=130 kDa;  
 CC IsoId=Q06190-1; Sequence=Displayed;  
 CC Name=PR72; Synonyms=72 kDa;  
 CC IsoId=Q06190-2; Sequence=VSP\_005107, VSP\_005108;  
 CC -1- TISSUE SPECIFICITY: Expressed in heart, brain, placenta, lung,  
 muscle and kidney.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L07590; AAB02613.1; -;  
 DR EMBL: L12146; AAB02614.1; -;  
 DR EMBL: AL389975; CAB97532.1; -;  
 DR PIR: A47114; A47114.  
 DR PIR: B47114; B47114.  
 DR Genew: HGNC:9307; PPP2R3A.  
 DR MIM: 604944; -;  
 DR GO: GO:0008601; P:protein phosphatase type 2A, intrinsic regu. . .; TAS.  
 DR InterPro: IPR002048; EF-hand.  
 DR Pfam: PF00036; efhand. 1.  
 DR PROSITE: PS00018; EF\_HAND; 1.  
 KW Calcium-binding; Multigene family; Alternative splicing.  
 FT DOMAIN 670 685 PRO-RICH  
 FT CA BIND 985 996 EF-HAND (POTENTIAL).  
 FT VARSPLIC 1 621 Missing (in isoform PR72).  
 FT VARSPLIC 622 665 /FTId=VSP\_005107.  
 FT MOIETLTSSQANLVCVRSPVGDKAKDTTSAAVLIOOTPE  
 FT VKK -> MWIKETSLRDPDLGRLAFLARGDFVLPSPRFK  
 FT KALKSPQQTQ (in isoform PR72).  
 FT /FTId=VSP\_005108.  
 SQ SEQUENCE 1150 AA; 130277 MW; 97A31BA4206518A3 CRC64;  
 Query Match 4.5%; Score 105.5; DB 1; Length 1150;  
 Best Local Similarity 19.9%; Pred. No. 17; Indels 109; Gaps 24;  
 Matches 99; Conservative 85; Mismatches 205;  
 25 GNKVVLGKKGTVE-----LTCTASQKKSIOFHWKNSNQ-----IKITLNGQS-- 67

Db 259 GSSTSESGNDTSSSETVYVNWVTRLASYLKLPFEFMQSGNNALDLTELISNWSLQ 318  
 QY 68 -----FLTKSPSKLNDRAKSRSLMDQGNFPI-IRKLNK-----IDSOT-- 106  
 Db 319 LTPSPVFGTEQPPKXYDVVQLSAS--DSGRFQITELQDNKPNRKKMDYVQSINNTNS 376  
 QY 107 -YIEVEVDOK--EYVQLVFLGTAN-----SDTHLQGSITLTLSPSGSPVQCRSP 158  
 Db 377 LYNLEAVNDPRTKRAVQVQSILTNPLNVSDDLME-----TLYIEESDGKALD---- 428  
 QY 159 RGNKIQGG---KTLVSQ--LELDQSGTW--TC-TVLQNKVKEFKIDIVVLAFOKASSI 210  
 Db 429 KGQKTENGPSHELLKVNHEHRAEPFEHATHLKCTPTMQNEIGKIFKSPVNLPREDCSK 488  
 QY 211 VYK-KEEGVYEFSPPLA-FIVEKLTSGGELMWQ-----AERASSKSWITFDLKNKVS 262  
 Db 489 VSKPEBDQDQDFNSSQGEIDIKLMDLESFSQMETSLREPLAKGNSNPLNSHSLQVTG 548  
 QY 263 VKRVTQDPKQLQMGKULPLHLTLPOALPOVAGSGLTLALSAKTKLHQEVNLVYMRATQL 322  
 Db 549 QTLVLEPKSKVSS--PIEKVSPG-----LTRIIFNGKIEEDRALLRILLES 597  
 QY 323 QKNLTCEWGGPTSEPKMLSLKENKEKAVSKREKPVNL-----NPEAGMQC----- 370  
 Db 598 IEDPAQEL-----VECKSRGSLQEKEMQILQETLTTSSQANLVCVRSPVGD 646  
 QY 371 -LLSDSQVLLSENIKYLPTWSTVPHRASALRPPIPGSALPDPDQTASALPDPPAASALP 429  
 Db 647 KAKDTTSAVLIQPTPEIKIKQNKREKKGTPLPAPTSPSSPRP--LSPVHNNVNVAP 704  
 QY 430 AALAVISFLGLGLGVAC 447  
 Db 705 LSNIRPFYFPEGGLPDTG 722  
 RESULT 82  
 ID 112B BOVIN STANDARD; PRT; 327 AA.  
 AC P46282;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Interleukin-12 beta chain precursor (IL-12B) (Cytotoxic lymphocyte  
 maturation factor 40 kDa subunit) (CLMF p40).  
 GN IL12B.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OX Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Holstein; TISSUE=Lymphoid;  
 RX MEDLINE=95244628; PubMed=7727547;  
 RA Zarlenaga D.S., Canals A., Aschenbrenner R.A., Gasbarré L.C.;  
 RT "Enzymatic amplification and molecular cloning of cDNA encoding the  
 small and large subunits of bovine interleukin 12."  
 RL Biochim. Biophys. Acta 1270:215-217(1995).  
 CC -1- FUNCTION: Cytokine that can act as a growth factor for activated T  
 CC and NK cells, enhance the lytic activity of NK/lymphokine-  
 CC activated killer cells, and stimulate the production of IFN-gamma  
 CC by resting PBMC (by similarity).  
 CC -1- SUBUNIT: Disulfide-bonded heterodimer of 40 kDa and 35 kDa  
 CC subunits.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- SIMILARITY: Belongs to the type I cytokine family of receptors.  
 CC Subfamily 3.  
 CC -1- SIMILARITY: Contains 1 fibronectin type III domain.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, U1815; AAA85792.1; -.
DR GO; GO:0005615; C:extracellular space; NMS.
DR GO; GO:0005143; F:interleukin-12 receptor binding; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0030546; F:receptor activator activity; ISS.
DR GO; GO:0042095; P:interferon-gamma biosynthesis; IDA.
DR GO; GO:0030101; P:natural killer cell activation; ISS.
DR GO; GO:0042104; P:positive regulation of activated T-cell pro. .; IDA.
DR GO; GO:0042035; P:regulation of cytokine biosynthesis; ISS.
DR GO; GO:0042093; P:T-helper cell differentiation; ISS.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003530; Hemocoptn_L_F3.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_C2.
DR Pfam; PF00041; Fn3; 1.
DR SMART; SM0060; FN3; 1.
DR SMART; SM00408; IGC2; 1.
DR PROSITE; PS01354; HEMATOPO_REC_L_F3; 1.
DR PROSITE; PS0835; IG_LIKE; 1.
DR Cytokine; Glycoprotein; Immunoglobulin domain; Signal.
FT SIGNAL 1 22
FT CHAIN 1 227
FT DOMAIN 23 327 INTERLEUKIN-12 BETA CHAIN.
FT DOMAIN 236 319 IG-LIKE C2-TYPE.
FT DISULFID 50 90 FIBROBLASTIN TYPE-III.
FT DISULFID 200 200 POTENTIAL.
FT CARBOHYD 223 223 INTERCHALIN (PROBABLE).
SQ SEQUENCE 327 AA; 37050 MW; D484DA36E6F3D9F CRC64;

Query Match 4.5%; Score 105; DB 1; Length 327;
Best Local Similarity 25.0%; Pred. No. 3.7; Indels 70; Gaps 17;
Matches 69; Conservative 38; Mismatches 99;

Qy 136 GOSLTLLTSPSPSPVOCSPRGKNIQG-KTILSVSLLELDSDGTCTVLONQKVE 194
Db 43 GETVLTCTDP--EEDGITWTSQSSSEVLGSGKTLTIQVKEGDAQNYTC-----HKGE 95

Qy 195 FKIDIVLTAFOKASSI----VYKGEQVEFSFPFLATYTKTGTSGELMWQERASSKKS 250
Db 96 ALSRSLLHLHKEDGIWSTDILKDQKEPKAKSF-LKCEAKDYSGHFTCWLTA-----147

Qy 251 WITFDLKNKEVSK--RVYQDPKLQWKKLPLHLTLPLQALPOYAGSGNLTLLAEKTKGL 308
Db 148 -ISTDLK--FSYKSRSGSSDPR-----GVTCGAL-----LSAE-KVSLR 183

Qy 309 HOEVNLVWRATQLQKLTCEVWGPTSPKMLT-----SLKLENKAKYSKRE--KP 357
Db 184 HREYNKYTV--ECQGSAC-----PALESLLLEVVEAVHKLKTYNTSSPFIDIIKP 236

Qy 358 VVVLNPEAGMOCCLSDSGQVLIESIKVLPFTWSTP 393
Db 237 ----DPPKNIQLRPLKNSRQV--EVSWEYDPDTWSTP 266

RESULT 83
OPCM_HUMAN STANDARD; PRT; 345 AA.
AC 014982;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Opioid binding protein/cell adhesion molecule precursor (OBSCAM)
DE (Opioid-binding cell adhesion molecule) (OPCML).
GN OPCML OR OBSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=occipital cortex;
RX MEDLINE=95237612; PubMed=7721093;
RA Shark K.B., Lee N.M.;
RT "Cloning, sequencing and localization to chromosome 11 of a CDNA
RT encoding a human opioid-binding cell adhesion molecule (OBSCAM).";
RL Gene 15:213-217(1995).
CC -1- FUNCTION: Binds opiate in the presence of acidic lipids; probably
CC involved in cell contact.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC similarity).
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. IgLN
CC family.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, U34774; AAA36387.1; -.
DR PIR, JC4025; JC4025.
DR Genew; HGNC:8143; OPCML.
DR MIM; 600632; -.
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR GO; GO:0004985; F:opiate receptor activity; TAS.
DR GO; GO:0007155; F:cell adhesion; TAS.
DR GO; GO:0008038; P:neutrophil cell recognition; TAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_C2.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00408; IGC2; 2.
DR PROSITE; PS0835; IG_LIKE; 3.
DR Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW Repeat; Signal; Lipoprotein.
FT SIGNAL 1 27
FT CHAIN 28 322
FT PROPEP 323 345
FT DOMAIN 39 126
FT DOMAIN 136 219
FT DOMAIN 223 310
FT DISULFID 57 115
FT DISULFID 157 202
FT DISULFID 244 296
FT CARBOHYD 44 44
FT CARBOHYD 70 70
FT CARBOHYD 140 140
FT CARBOHYD 285 285
FT CARBOHYD 293 293
FT CARBOHYD 306 306
FT LIPID 322 322
SQ SEQUENCE 345 AA; 38007 MW; E7AD17BEA1AA3F4 CRC64;

Query Match 4.5%; Score 105; DB 1; Length 345;
Best Local Similarity 21.3%; Pred. No. 3.9; Indels 54; Gaps 8;
Matches 48; Conservative 40; Mismatches 83;

Qy 11 LVLQTLALPAATQGNKVLG-----KKGPTVELCTASQKSIQFMKSN 57
Db 14 LVVVSRLILFLVPTVPVPRSGDATPPKAMDVTVYQSGSATLRCTIDRVT-RVAMLRKS 72

Qy 58 QIKILGN-----QGSFLTKGPKSLNDRADSRSLWDQGNFPLIKLKIEDSDTYICE 110
Db 73 TLVAGNDKWSIDPVIILVTVPTQ-----YSIMIQVADVYDEGPTCS 116
```





FT	CARBOHYD	287	287	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	351	351	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	CARBOHYD	398	398	N-LINKED (GLCNAC. . .) (POTENTIAL)
FT	SEQUENCE	515 AA;	58445 MW;	ICBB9D5A024A41E8 CRC64;
Query Match		4.5%	Score 104.5;	DB 1; Length 515;
Best Local Similarity		20.1%	Pred. No. 7.2;	Indels 103; Gaps 17
Matches		68;	Conservative 48;	Mismatch 120;
QY	168	TLVSQLEL-QDSGTCTV-LONO-----KVEFKIDI-VILAFOKASSIVYKKEG	217	
DB	19	SLTLLTLLALCGPIQWRCSLSLGNQOMWTMAYNGAKFSSIDQILLEANHNSFFCAKSP--	76	
QY	218	QVESFPLAFVYEKLTGSGELMW---QERASSSKSWITEDIKNEVSKVRYTODPKLOM	274	
DB	77	-----RYTLDSVNGYPKIYWPDPQGRRRFGADAMVYD-----EPRC--	114	
QY	275	GKULPLHLTPALQOYAGSGLT-----ALEATGKLNQEVNLVWKRATQLOK--NLT	327	
DB	115	-----PYVGADRFDCPHMDNASQAQSGSYFNHQILFLHKKOCHGIFTLT	159	
QY	328	CEVMGPTSPKMLSL-KLENKEAKYSKREKPVNLVPEAKMOCLLSDGQVLTESNI--	384	
DB	160	WEING-YDPLIRFSLHKIPDPQDPFQPLNSDW--PSVRSALLNLQYARAFDPCALCW	216	
QY	385	KVLPTWSTFVPHRASALPAPPTGSLPDPQ-----TAS	417	
DB	217	EPSPMABEILVYNKTISSSGGLALPDQIIFWNTSSEPTTQGMHHPGQRLLFNVSQGN	276	
QY	418	ALPDP-----AASALPAL---AVISFLIGLGLV	445	
DB	277	ALLPPLSLVNLSTASSAPPTVRRSVVALTLGLALSV	315	
RESULT 86				
ECTO_RAT		STANDARD;	PRT;	519 AA.
ID	ECTO_RAT			
AC	P16573;			
DT	01-AUG-1990 (Rel. 15, Created)			
DT	01-OCT-1994 (Rel. 30, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Ecto-ATPase precursor (Cell-CAM 105) (C-CAM 105) (ATP-dependent			
DE	taurocolate-carrier protein) (gp110).			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxId:10116;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORM LONG), AND SEQUENCE OF 50-68.			
RC	STRAIN=Spregue-Dawley; TISSUE=Liver;			
RX	MEDLINE=89340561; PubMed=2557225;			
RA	Lin S.-H., Guidotti G.;			
RT	"Cloning and expression of a cDNA coding for a rat liver plasma			
RT	membrane ecto-ATPase. The primary structure of the ecto-ATPase is			
RT	similar to that of the human biliary glycoprotein I.";			
RL	J. Biol. Chem. 264:14408-14414(1989).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM SHORT).			
RC	STRAIN=Spregue-Dawley, and Miscar; TISSUE=Liver;			
RX	MEDLINE=93379310; PubMed=8504806;			
RA	Edlund M., Gaardsvoll H., Bock E., Oebirnk B.;			
RT	"different isoforms and stock-specific variants of the cell adhesion			
RT	molecule C-CAM (cell-CAM 105) in rat liver.";			
RL	Eur. J. Biochem. 213:1109-1116(1993).			
RN	[3]			
RP	SEQUENCE FROM N.A. (ISOFORM SHORT).			
RC	STRAIN=Spregue-Dawley; TISSUE=Liver;			
RX	MEDLINE=92344597; PubMed=1637321;			
RA	Culic O., Huang O., Flanagan D., Hixon D., Lin S.-H.;			
RT	"Molecular cloning and expression of a new rat liver cell-CAM105			
RT	isoform. Differential phosphorylation of isoforms.";			
RL	Biochem. J. 285:47-53(1992).			
RN	[4]			

```

RP SEQUENCE FROM N.A..
RC TISSUE=Intestine; PubMed=8240240;
RX MEDLINE=9405980;
RA Cheung P.H., Culic O., Qiu Y., Earley K., Thompson N., Hixson D.C.,
RL Lin S.-H.;
RT "type cytoplasmic domain of C-CAM is required for C-CAM-mediated
RT adhesion function: studies of a C-CAM transcript containing an
RT unspliced intron."/;
RL Biochem. J. 295:427-435(1993).
[5]
RP PARTIAL SEQUENCE.
RX MEDLINE=90292222; PubMed=2145577;
RA Aurtivillius M., Hansen O.C., Lazrek M.B.S., Bock E., Oebink B.;
RT "The cell adhesion molecule Cell-CAM 105 is an ecto-ATPase and a
RT member of the immunoglobulin superfamily.";
RL FEBS Lett. 264:267-269(1990).
[6]
RP PARTIAL SEQUENCE.
RC STRAIN=Miscar; TISSUE=Liver;
RX MEDLINE=93292517; PubMed=8513803;
RA Becker A., Lucka L., Kilian C., Kamnitsch C., Reuter W.;
RT "Characterisation of the ATP-dependent taurocholate-carrier protein
RT (gp110) of the hepatocyte canalicular membrane."/;
RL Eur. J. Biochem. 214:539-548(1993).
[7]
RP ALTERNATIVE SPLICING.
RX MEDLINE=93123234; PubMed=8380406;
RA Najjar S.M., Accilli D., Philippe N., Jernberg J., Margolis R.,
RA Taylor S.I.;
RT "p120/ecto-ATPase, an endogenous substrate of the insulin receptor
RT tyrosine kinase, is expressed as two variably spliced isoforms."/;
RL J. Biol. Chem. 268:1201-1206(1993).
[8]
RP CHARACTERIZATION.
RX MEDLINE=91354197; PubMed=1831973;
RA Lin S.-H., Culic O., Flanagan D., Hixson D.C.;
RT "Immunohistochemical characterization of two isoforms of rat liver ecto-
RT ATPase that show an immunological and structural identity with a
RT glycoprotein cell-adhesion molecule with Mr 105,000."/;
RL Biochem. J. 278:155-161(1991).
-1- FUNCTION: THIS PROTEIN IS A CALCIUM-INDEPENDENT CELL ADHESION
MOLECULE WITH HOMOPHILIC BINDING PROPERTIES. MAY PLAY A ROLE IN
THE FORMATION OF THE APICAL SURFACE OF THE SPECIALIZED MEMBRANE
STRUCTURE OF THE APICAL SURFACE OF THE HEPATOCYTES.
-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CANALICULAR DOMAIN
OF HEPATOCYTE PLASMA MEMBRANS.
-1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=Long;
IsoId=p16573-1; Sequence=Displayed;
Name=Short; Synonyms=C-CAM3;
IsoId=p16573-2; Sequence=VSP_002504;
-1- TISSUE SPECIFICITY: EXPRESSED IN EPITHELIA, VESSEL ENDOTHELIA,
LEUKOCYTES AND PLATELETS.
-1- PTM: PHOSPHORYLATED ON SERINE RESIDUES IN BOTH ISOFORMS. ON SERINE
AND TYROSINE RESIDUES IN CYTOPLASMIC REGION OF THE LONG ISOFORM.
-1- SIMILARITY: Belongs to the immunoglobulin superfamily. CEA family.
-1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
-1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
-----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/isb-ch).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04963; AAA41104.1; -
DR EMBL; X71122; CAA50435.1; -
DR EMBL; Z12019; CAA78054.1; -
DR EMBL; M92848; AAA16783.1; ALT_TERM.
DR PIR; A44783; A44783.

```

[illegible]

Query Match	4.5%	Score 104.5;	DB 1;	Length 519;
Best Local Similarity	19.0%;	Pred. No. 7.2;		
Matches 99;	Conservative 84;	Mismatches 177;	Indels 161;	Gaps 27;

```

QY 3 RG-VFRRHLLVLQ-----ALPAAQGNKVLQKGDYELFCTAQR 46
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 10 RQGFPMRGLLLTASLLTYWSPLLTAQTVDAVPPRVVESSVLL-----LTHNLRQE 61
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 47 KSIQEHMKNQSIKILNQGSGFLTGPBKLNDRAISR-SLMDQGNFLIILKILIEDSD 105
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 62 FQV-FPMYKVTTTGAINSEIARIYRSNSTDQEPAPSGRYTYSNQS--LFPNVAKTBEG 118
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 106 TYICEVEDQ-----KEEVOQLVF-----GLTANSDLHLLQGSLLTLTLESPGSSPSVQ 154
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 119 PYTLSDVLDKQENPIDTSGVOFRVYPAQLQKNVTGN--NSNPMDEEFP--VALMCEPPYNNSTY 176
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

Oy 155 CREPRKKN-----OQGTTLSVSGLELSDSGTWTCTVLQ---QKRYEFKIDIV- 200
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 177 LWSRNESESLEBDRTYFSEGRNRTLLNLNVRTDKGYECBEARNATNRRDPRFLVDIYG 236
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 201 ---VLAFOKASSIYKKBGEQVEFSF-----PLA---FTVEKL-TGSGELWQMERAS 246
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 237 PDAPVI-----SPPDYILHQGSMNLNLSCHADSNPFAOYFWMLNKLQTSOELF.SNITT 292
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 247 SSKSWITTFDLK-----KEVSRYKVIQDPKLCQKGLPLHLTLPLPALPOYAGSGN 296
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 293 NSGTYACF-VNNTVTGLSTRTYKNTITVEPTQSIQTNTTYKEL-----GS 339
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 297 LTLALSAK-TGKLHQEVALVYMRATOLQKLTCEYWGPTSPKML-----SIKLENKAKY 351
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 340 VTLTCESKDTG-----VSVR-----WLFNSQSLDTRMLTSDONSTLRI 379
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 352 S--KREKPVWVLPENKGMQCLLSDSGVLLNESNIKVLPTSTFVHPRASALPAPPTGSA 409
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 380 DPIKRE-----DAGDQCEISN-----PVSFR 401
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 410 LPDPQTASALPPPPAASALPALALVISTFLGLGAGVACVLA 450
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 402 ISHPILKLDVIPDTQSGNSGLAGIAGIVISVAGVAILAA 442
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 87
PTK7_HUMAN
ID_PTK7_HUMAN STANDARD; PRT: 1070 AA.
AC_Q13306; Q13417;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Tyrosine-protein kinase-like 7 precursor (Colon carcinoma kinase-4)
DE (CK4).
GN PTK7 OR CK4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
[1]
[2]
[3]
[4]
RA Park S.-K., Lee H.-S., Lee S.-T.;
RT "Characterization of the human full-length PTK7 cDNA encoding a
RT receptor protein tyrosine kinase-like molecule closely related to
RT chick KLG."
RL J. Biochem. 119:235-239 (1996).
[5]
[6]
[7]
[8]
[9]
[10]
[11]
[12]
[13]
[14]
[15]
[16]
[17]
[18]
[19]
[20]
[21]
[22]
[23]
[24]
[25]
[26]
[27]
[28]
[29]
[30]
[31]
[32]
[33]
[34]
[35]
[36]
[37]
[38]
[39]
[40]
[41]
[42]
[43]
[44]
[45]
[46]
[47]
[48]
[49]
[50]
[51]
[52]
[53]
[54]
[55]
[56]
[57]
[58]
[59]
[60]
[61]
[62]
[63]
[64]
[65]
[66]
[67]
[68]
[69]
[70]
[71]
[72]
[73]
[74]
[75]
[76]
[77]
[78]
[79]
[80]
[81]
[82]
[83]
[84]
[85]
[86]
[87]
[88]
[89]
[90]
[91]
[92]
[93]
[94]
[95]
[96]
[97]
[98]
[99]
[100]
[101]
[102]
[103]
[104]
[105]
[106]
[107]
[108]
[109]
[110]
[111]
[112]
[113]
[114]
[115]
[116]
[117]
[118]
[119]
[120]
[121]
[122]
[123]
[124]
[125]
[126]
[127]
[128]
[129]
[130]
[131]
[132]
[133]
[134]
[135]
[136]
[137]
[138]
[139]
[140]
[141]
[142]
[143]
[144]
[145]
[146]
[147]
[148]
[149]
[150]
[151]
[152]
[153]
[154]
[155]
[156]
[157]
[158]
[159]
[160]
[161]
[162]
[163]
[164]
[165]
[166]
[167]
[168]
[169]
[170]
[171]
[172]
[173]
[174]
[175]
[176]
[177]
[178]
[179]
[180]
[181]
[182]
[183]
[184]
[185]
[186]
[187]
[188]
[189]
[190]
[191]
[192]
[193]
[194]
[195]
[196]
[197]
[198]
[199]
[200]
[201]
[202]
[203]
[204]
[205]
[206]
[207]
[208]
[209]
[210]
[211]
[212]
[213]
[214]
[215]
[216]
[217]
[218]
[219]
[220]
[221]
[222]
[223]
[224]
[225]
[226]
[227]
[228]
[229]
[230]
[231]
[232]
[233]
[234]
[235]
[236]
[237]
[238]
[239]
[240]
[241]
[242]
[243]
[244]
[245]
[246]
[247]
[248]
[249]
[250]
[251]
[252]
[253]
[254]
[255]
[256]
[257]
[258]
[259]
[260]
[261]
[262]
[263]
[264]
[265]
[266]
[267]
[268]
[269]
[270]
[271]
[272]
[273]
[274]
[275]
[276]
[277]
[278]
[279]
[280]
[281]
[282]
[283]
[284]
[285]
[286]
[287]
[288]
[289]
[290]
[291]
[292]
[293]
[294]
[295]
[296]
[297]
[298]
[299]
[300]
[301]
[302]
[303]
[304]
[305]
[306]
[307]
[308]
[309]
[310]
[311]
[312]
[313]
[314]
[315]
[316]
[317]
[318]
[319]
[320]
[321]
[322]
[323]
[324]
[325]
[326]
[327]
[328]
[329]
[330]
[331]
[332]
[333]
[334]
[335]
[336]
[337]
[338]
[339]
[340]
[341]
[342]
[343]
[344]
[345]
[346]
[347]
[348]
[349]
[350]
[351]
[352]
[353]
[354]
[355]
[356]
[357]
[358]
[359]
[360]
[361]
[362]
[363]
[364]
[365]
[366]
[367]
[368]
[369]
[370]
[371]
[372]
[373]
[374]
[375]
[376]
[377]
[378]
[379]
[380]
[381]
[382]
[383]
[384]
[385]
[386]
[387]
[388]
[389]
[390]
[391]
[392]
[393]
[394]
[395]
[396]
[397]
[398]
[399]
[400]
[401]
[402]
[403]
[404]
[405]
[406]
[407]
[408]
[409]
[410]
[411]
[412]
[413]
[414]
[415]
[416]
[417]
[418]
[419]
[420]
[421]
[422]
[423]
[424]
[425]
[426]
[427]
[428]
[429]
[430]
[431]
[432]
[433]
[434]
[435]
[436]
[437]
[438]
[439]
[440]
[441]
[442]
[443]
[444]
[445]
[446]
[447]
[448]
[449]
[450]
[451]
[452]
[453]
[454]
[455]
[456]
[457]
[458]
[459]
[460]
[461]
[462]
[463]
[464]
[465]
[466]
[467]
[468]
[469]
[470]
[471]
[472]
[473]
[474]
[475]
[476]
[477]
[478]
[479]
[480]
[481]
[482]
[483]
[484]
[485]
[486]
[487]
[488]
[489]
[490]
[491]
[492]
[493]
[494]
[495]
[496]
[497]
[498]
[499]
[500]
[501]
[502]
[503]
[504]
[505]
[506]
[507]
[508]
[509]
[510]
[511]
[512]
[513]
[514]
[515]
[516]
[517]
[518]
[519]
[520]
[521]
[522]
[523]
[524]
[525]
[526]
[527]
[528]
[529]
[530]
[531]
[532]
[533]
[534]
[535]
[536]
[537]
[538]
[539]
[540]
[541]
[542]
[543]
[544]
[545]
[546]
[547]
[548]
[549]
[550]
[551]
[552]
[553]
[554]
[555]
[556]
[557]
[558]
[559]
[560]
[561]
[562]
[563]
[564]
[565]
[566]
[567]
[568]
[569]
[570]
[571]
[572]
[573]
[574]
[575]
[576]
[577]
[578]
[579]
[580]
[581]
[582]
[583]
[584]
[585]
[586]
[587]
[588]
[589]
[590]
[591]
[592]
[593]
[594]
[595]
[596]
[597]
[598]
[599]
[600]
[601]
[602]
[603]
[604]
[605]
[606]
[607]
[608]
[609]
[610]
[611]
[612]
[613]
[614]
[615]
[616]
[617]
[618]
[619]
[620]
[621]
[622]
[623]
[624]
[625]
[626]
[627]
[628]
[629]
[630]
[631]
[632]
[633]
[634]
[635]
[636]
[637]
[638]
[639]
[640]
[641]
[642]
[643]
[644]
[645]
[646]
[647]
[648]
[649]
[650]
[65
```

```

CC KIDNEY, PLACENTA AND MELANOCYTES. WEAKLY EXPRESSED IN THYROID
CC GLAND, OVARY, BRAIN, HEART AND SKELETAL MUSCLE. ALSO EXPRESSED IN
CC ERYTHROCYTES AND CELLS. BUT NOT EXPRESSED IN COLON.
CC -1- SIMILARITY: Belongs to the Tyr family of protein kinases. Insulin
CC receptor subfamily.
CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, U33635; AAA87565.1; -
DR EMBL, U40271; AAC50484.2; -
DR EMBL, AF447176; AAL39062.1; -
DR EMBL, AF447157; AAL39062.1; JOINED.
DR EMBL, AF447158; AAL39062.1; JOINED.
DR EMBL, AF447162; AAL39062.1; JOINED.
DR EMBL, AF447164; AAL39062.1; JOINED.
DR EMBL, AF447167; AAL39062.1; JOINED.
DR EMBL, AF447170; AAL39062.1; JOINED.
DR EMBL, AF447171; AAL39062.1; JOINED.
DR EMBL, AF447173; AAL39062.1; JOINED.
DR EMBL, AF447174; AAL39062.1; JOINED.
DR EMBL, AF447175; AAL39062.1; JOINED.
DR HSP, P08631; IAD5.
DR Genew; HGNC:9618; PKT7.
DR MIM; 601890; -.
DR GO; GO:0005886; C:plasma membrane; TAS.
DR GO; GO:0005888; C:proteoglycan integral to plasma membrane; TAS.
DR GO; GO:0004714; F:transmembrane receptor protein tyrosine kin. .; TAS.
DR GO; GO:0007165; P:signal transduction; TAS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG-C2.
DR InterPro; IPR003006; IG-MHC.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002011; ReceptorTyrosineK.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR008266; Tyr_kinase_AS.
DR Pfam; PF00047; Ig_7.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00408; IGC2; 5.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PSS0835; IG_LIKE; 7.
DR PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PSS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PSS00239; RECEPTOR_TYR_KIN_II; FALSE_NEG.
KW Receptor; Transmembrane; Signal; Glycoprotein; Cell adhesion;
KW Immunoglobulin domain; Repeat.
FT CHAIN 1 1070 POTENTIAL.
FT SIGNAL 30
FT DOMAIN 31 704 TYROSINE-PROTEIN KINASE-LIKE 7.
FT TRANSMEM 705 725 EXTRACELLULAR (POTENTIAL).
FT POTENTIAL 726 1070
FT CYTOPLASMIC (POTENTIAL).
FT DOMAIN 31 120 IG-LIKE C2-TYPE 1.
FT DOMAIN 128 218 IG-LIKE C2-TYPE 2.
FT DOMAIN 225 317 IG-LIKE C2-TYPE 3.
FT DOMAIN 309 407 IG-LIKE C2-TYPE 4.
FT DOMAIN 412 497 IG-LIKE C2-TYPE 5.
FT DOMAIN 503 586 IG-LIKE C2-TYPE 6.
FT DOMAIN 578 680 IG-LIKE C2-TYPE 7.
FT DOMAIN 796 1066 PROTEIN KINASE; INACTIVE.
FT DISULFID 53 101 BY SIMILARITY.
FT DISULFID 150 200 BY SIMILARITY.
FT DISULFID 246 301 BY SIMILARITY.
FT DISULFID 343 391 BY SIMILARITY.
FT DISULFID 433 481 BY SIMILARITY.
FT DISULFID 524 570 BY SIMILARITY.

```

```

FT DISULFID 613 664 BY SIMILARITY.
FT CARBOHYD 116 116 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 283 283 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 405 405 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 463 463 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 567 567 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 92 92 P -> R (IN REF. 2 AND 3).
FT CONFLICT 147 147 K -> T (IN REF. 2 AND 3).
FT CONFLICT 207 207 S -> G (IN REF. 2 AND 3).
FT CONFLICT 495 496 VL -> RV (IN REF. 2 AND 3).
FT CONFLICT 515 515 G -> E (IN REF. 2 AND 3).
FT CONFLICT 881 881 E -> G (IN REF. 2 AND 3).
FT CONFLICT 969 969 A -> P (IN REF. 2 AND 3).
FT CONFLICT 992 992 S -> F (IN REF. 2 AND 3).
SQ SEQUENCE 1070 AA; 118260 MW; 47CDF25B8E3698A5 CRC64;

Query Match 4.5%; Score 104.5; DB 1; Length 1070;
Best Local Similarity 20.8%; Pred. No. 18;
Matches 86; Conservative 58; Mismatches 147; Indels 123; Gaps 22;

QY 89 QGNPFLIKKNIKIEDSDTYICEVED---OKEEVOLV---FGLTANSDFHLQGSGLT 140
DB 371 QKGHELVIANIAESDAGVYTHAANLAGORQDVIVATVPMSLKKRQDSOLEBGRFGY 430
QY 141 LTLSPSSSPSVQ-----CRSPRGKNIQGGKTLVSQLELDSTG-TG----- 184
DB 431 LDCLQATPKPTVWYRNQMLISEDSREVEFKNG-TLRINSVEYVD-CTWRKMSSTPAG 488
QY 185 -----TVLQNKQVEF---KIDIVLAFQKASSIYKKEGQVEFPLATVTEKLTSG 236
DB 489 SIEQAVLQVELEKLFTEPPQPOQCMGPEKEATVPSCATRG-----EKPT--- 533
QY 237 ELMQAEASASSKSKITFDLNKKEVSRYVQDPFLQMGKPLHLTLPLQALPVYAGSGN 296
DB 534 -IKWRADGSSIPENVYTNAGT--LHFAKVRD-----DIGN 567
QY 297 LT-LALEAKTKLHGEVNLVVM-----RATLOKN---LTCEVWGFTSPRLMTSL 342
DB 568 YTCIASNPOQGIARHVLAVAVITFKVEPERTTVYGTHALLQCEAGQ--DPRKLQW 625
QY 343 KLENKEAVSKREKFWVLN-----PE-AGMOCCLISDSQVLLESNIKYLPWIS 391
DB 626 KGDRIIDPTKGPBMHIFONGSLIVHDVAPEDSGRYTCIAGNS-----CNIK---HTE 676
QY 392 TPVHRAALPAPPTGSLPDP---QTASALPDPASALPALAVISFILGL 441
DB 677 APLVYVDPKVEBSESGSPPPYKMTITIGL-----SVGAAYIIVAVGL 722

RESULT 86
E2F2 HUMAN STANDARD; PRT; 437 AA.
ID E2F2_HUMAN
AC Q14209;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Transcription factor E2F2 (E2F-2).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94067141; PubMed=8246995;
RA Ivey-Hoyte M., Conroy R., Huber H.E., Goodhart P.J., Olfelt A.,
RA Heimbrock D.C.;
"Cloning and characterization of E2F-2, a novel protein with the

```

RT biochemical properties of transcription factor E2F." ;  
 RL Mol. Cell. Biol. 13:7802-7812(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Rieder M.J., Livingston R.J., Braun A.C., Montoya M.A., Chung M.-W.,  
 RA Miyamoto K.E., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,  
 RA Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;  
 RA Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Pearce A.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Transcription activator that binds DNA cooperatively  
 CC with DP proteins through the E2 recognition site, TTTCC/GCGC,  
 CC found in the promoter region of a number of genes whose products  
 CC are involved in cell cycle regulation or in DNA replication. The  
 CC DRF1/E2F complex functions in the control of cell-cycle  
 CC progression from G1 to S phase. E2F-2 binds specifically to Rb1  
 CC protein, in a cell-cycle dependent manner.  
 CC -1- SUBUNIT: Component of the DRF1/E2F transcription factor complex.  
 CC Forms heterodimers with DP family members. The E2F-2 complex binds  
 CC specifically hypophosphorylated retinoblastoma protein Rb1. During  
 CC the cell cycle, Rb1 becomes phosphorylated in mid-to-late G1  
 CC phase, detaches from the DRF1/E2F complex, rendering E2F  
 CC transcriptionally active. Viral oncoproteins, notably E1A, T-  
 CC antigen and HPV E7, are capable of sequestering Rb protein, thus  
 CC releasing the active complex.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: Highest level of expression is found in  
 CC placenta, low levels are found in lung. Found as well in many  
 CC immortalized cell lines derived from tumor samples.  
 CC -1- PTM: Phosphorylated by CDK2 and cyclin A-CDK2 in the S-phase (by  
 CC similarity).  
 CC -1- SIMILARITY: Belongs to the E2F/DP family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC DR EMBL: L22846; AAA16890.1; -  
 CC DR EMBL: AF518877; AAM54044.1; -  
 CC DR EMBL: AL021154; CAA15949.1; -  
 CC DR PIR: A54595; A54595.  
 CC DR PDB: 1N4M; 07-JAN-03.  
 CC DR TRANSFAC: T01544; E2F2.  
 CC DR GeneW: HGNC:3114; E2F2.  
 CC DR GK: Q14209; -  
 CC DR MIM: 600426; -  
 CC DR GO: GO:0003677; F:DNA binding; TAS.  
 CC DR GO: GO:0005515; F:protein binding; TAS.  
 CC DR GO: GO:0003702; F:RNA polymerase II transcription factor acti. ; TAS.  
 CC DR GO: GO:0006367; F:transcription initiation from Pol II promoter; TAS.  
 CC DR InterPro: IPR003316; E2F TDP.  
 CC DR Pfam: PF02319; E2F TDP; 1.  
 CC DR Transcription regulation; Activator; DNA-binding; Nuclear protein;  
 CC phosphorylation; Cell cycle; 3D-structure.  
 CC KW Phosphorylation; Cell cycle; 3D-structure.  
 CC FT DOMAIN 65 105 CYCLIN A/CDK2 BINDING (POTENTIAL).  
 CC FT DNA BIND 107 196 POTENTIAL.  
 CC FT DOMAIN 155 176 LEUCINE-ZIPPER.  
 CC FT DOMAIN 160 196 DEF BOX.  
 CC FT DOMAIN 197 289 DIMERIZATION (POTENTIAL).  
 CC FT DOMAIN 359 437 TRANSACTIVATION (POTENTIAL).  
 CC FT DOMAIN 360 363 POLY-PRO.  
 CC FT DOMAIN 410 427 RETINOBLASTOMA PROTEIN ASSOCIATION  
 CC (POTENTIAL).  
 CC SQ SEQUENCE 437 AA; 47505 MW; 60541F4235507005 CRC64;  
 CC Query Match 4.5%; Score 104; DB 1; Length 437;  
 CC Best Local Similarity 22.3%; Pred. No. 6.3;

Matches 75; Conservative 48; Mismatches 145; Indels 68; Gaps 15;  
 QY 145 SPGGSPSVOC-----RSPGKNIQ---GGKTVSVSLQELDSDGTWCTVQLQNKKEVFK 196  
 Db 53 APFAAPGT-CLDATTGPGPQVRCPLPAGLPLKRLDLBGICRPVVPPEPTFGKICIR 111  
 QY 197 IDIVLAFOKASIVYKKEGQVFEFPLAFVTEKL-----TSGELMWAQERASSS 248  
 Db 112 VDGL-----PSPTPSPSPKTRVDTSLGLTFKKFYLLSESDVLDLMAAEVDVQ 165  
 QY 249 KSWITPDLKKEVSVKVT-----QDP-----KLQMGKPLPHLTLPQAL 288  
 Db 166 KRRI-YDITVTLSEGIQIRKAKNNIQVGRGMEDEPTRPGQOQLGELKELMNTQAL 224  
 QY 289 PQVGSNLTLL-ALAKTGLHGVNLVVMRAT-QLQKVLTCRVMGPTSPKL----- 338  
 Db 225 DQLIQSCLSKLHTEKANKRLAYVYQDRAVGNKEQIVAVKAPQTRLVPPDRT 284  
 QY 339 -MSLKLENKAKVSKREKVPVVLNPEAGMWOCILSDSGVLTESNIKVLPTWSTP---V 394  
 Db 285 DNLQIVLKSQTGPIEVYLCPEEVEGPDSPSEEPSPSTs--TLCPSPSAQPSSTDPDSIM 342  
 QY 395 HPRASALPAPPTGSALDPDTASALPPPPAASALPA 430  
 Db 343 EPTASSVPAP---APTPQAA---PPPPSLVPLEA 370  
 RESULT 89  
 BUTY MOUSE STANDARD; PRT; 524 AA.  
 AC Q62556; P97392;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Butyrophilin precursor (BT) (Butyrophilin subfamily 1 member A1).  
 GN BTN1A1 OR BTN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=129; TISSUE=Mammary gland;  
 RC MEDLINE=97148936; PubMed=8995761;  
 RA Ogg S.L., Komaragiri M.V.S., Mather I.H.;  
 RT "Structural organization and mammary-specific expression of the  
 RT butyrophilin gene." ;  
 RN Mamm. Genome 7:900-905(1996).  
 RN [2]  
 RP SEQUENCE OF 39-487 FROM N.A.  
 RC TISSUE=Mammary gland;  
 RX MEDLINE=96125722; PubMed=8541302;  
 RA Ishii T., Aoki N., Noda A., Adachi T., Nakamura R., Matsuda T.;  
 RT "Carboxy-terminal cytoplasmic domain of mouse butyrophilin  
 RT cells and milk fat globule membrane." ;  
 RL Biochem. Biophys. Acta 1245:285-292(1995).  
 CC -1- FUNCTION: May function in the secretion of milk-fat droplets. It  
 CC may act as a specific membrane-associated receptor for the  
 CC association of cytoplasmic droplets with the apical plasma  
 CC membrane (By similarity).  
 CC -1- SUBUNIT: Seems to associate with xanthine dehydrogenase/oxidase.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: Expressed in mammary tissue and secreted in  
 CC association with the milk-fat-globule membrane during lactation.  
 CC -1- DEVELOPMENTAL STAGE: Expression increases during the last half of  
 CC pregnancy and is maximal during lactation.  
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. BTN/MOG  
 CC family.  
 CC -1- SIMILARITY: Contains 2 immunoglobulin-like V-type domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC EMBL; U67065; AAB51034.1; -

DR EMBL; S80642; AAB55893.1; -

DR MGI; 103118; Btlna1.

DR InterPro; IPR001870; B302.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003596; Ig V.

DR InterPro; IPR006574; PRV.

DR InterPro; IPR003877; SPRY\_receptor.

DR Pfam; PF00622; SPRY; 1.

DR SMART; SM00406; IGV; 1.

DR SMART; SM00589; PRV; 1.

DR SMART; SM00449; SPRY; 1.

DR PROSITE; PS00835; IG LIKE; 2.

KM Transmembrane; Glycoprotein; Immunoglobulin domain; Signal; Repeat.

FT SIGNAL 1

FT CHAIN 26

FT DOMAIN 27 524

FT TRANSMEM 27 247

FT DOMAIN 248 268

FT DOMAIN 269 524

FT DOMAIN 29 141

FT DOMAIN 149 235

FT CARBOHYD 56 216

FT CARBOHYD 46 216

FT CONFLICT 117 117

FT CONFLICT 191 191

FT CONFLICT 210 210

FT CONFLICT 363 363

FT CONFLICT 408 408

FT CONFLICT 413 414

FT CONFLICT 420 423

FT CONFLICT 492 509

SEQ 524 AA; 58406 MW; 333FADE2C7704480 CRC64;

Query Match 4.5%; Score 104; DB 1; Length 524;

Best Local Similarity 22.5%; Pred. No. 7.9; Mismatches 108; Indels 92; Gaps 17;

Matches 73; Conservative 52; Mismatches 108; Indels 92; Gaps 17;

QY 11 LVLVQLALLPA-----TQGNKVVYGGKGDVVELCTASQKKSIOF---HMKNNSNOIKI 61

DB 16 LTVLQFLPLDPAFPDVTAPQEPVLAIVGSDALITCGSPVASEVWELLMFRQTRSTAV 75

QY 62 L-----GNQGSFLT--KGPSEKLNDRADSRSLMDQGNFLITKLIKIEDSDYICVEYD 113

DB 76 LLYRDGGEQGEQOMTEYRGRATL-----ATAGLID-GRATLLIDIVRSDGEYRCLFKD 129

QY 114 -----QKEEVLVFGLTANSPTHLQGGSLTLTLESPPGS---SPSVQCRS----- 157

DB 130 NDDEBAAVYIKVAVGSDDPOISMTVGENGEMIECTSSGGYRPPQVQWRKGNREMLPST 189

QY 158 --PRGNKIOGGKTLVSQLELDGSGTWTCTVLQNKQKVEFKIDIVLAFOKASSIVYKKE 215

DB 190 SESKKNHBEGLFTVAVSM-IRDSI-----KMSKCIQNILG-----Q 228

QY 216 GEQVEFSPFLAFTYEKLT-----GSGELMWQAEKASSKSKWTTDLK 257

DB 229 GKEVEISLPAPF-VPRILTPWIVAAVAILLALGFLTISIFPTWLYKERRS-----LR 280

QY 258 NKEV-SVKRVTPDPLQWKKPLH 281

DB 281 KKEFGSKERLLBEELRC---KKTUHL 302

RESULT 90

CD22\_MOUSE

ID CD22\_MOUSE STANDARD; PRT; 862 AA.

AC P31329; 09JH12; 09JHX9; 09JY10; 09JY11; 09R056; 09R094; 09WU51;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE B-cell receptor CD22 precursor (Ieu-14) (B-lymphocyte cell adhesion molecule) (BL-CAM) (Siglec-2).

GN CD22 OR LYB-8.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_Taxid=10090;

QY [1]

RP SEQUENCE FROM N. A.

RC STRAIN=DRB/2J, and BALB/C. TISSUE=Liver;

RK MEDLINE=9315834; PubMed=8100843;

RA Law C., Torres R.M., Sundberg H.A., Parkhouse R.M., Brennan C.I., Copeland N.G., Jenkins N.A., Clark E.A.;

RT "Organization of the murine Cd22 locus. Mapping to chromosome 7 and characterization of two alleles.";

RL J. Immunol. 151:175-187(1993).

RL [2]

RP SEQUENCE FROM N. A.

RC STRAIN=C57BL/6, BXB, and MRL;

RK MEDLINE=99432012; PubMed=10501843;

RA Lajunias F., Inou-Zekri N., Fossati Jimack L., Chieportiche Y., Parkhouse R.M., Mary C., Reininger L., Brishouse G., Izui S.;

RL "Polymorphisms in the Cd22 gene of inbred mouse strains.";

RL Immunogenetics 49:991-995(1999).

RL [3]

RP SEQUENCE OF 1-178 FROM N. A. (ISOFORMS 1; 2 AND 3).

RC STRAIN=NZW; TISSUE=Spleen;

RK MEDLINE=20432275; PubMed=10975807;

RA Mary C., Laporte C., Parzy D., Santiago M.L., Stefani F., Lajunias F., Parkhouse M.E., O'Keefe T.L., Neuburger M.S., Izui S., Reininger L.;

RT "Dysregulated expression of the Cd22 gene as a result of a short interspersed nucleotide element insertion in Cd22alpha lupus-prone mice.";

RL J. Immunol. 165:2987-2996(2000).

RL [4]

RP SIALIC ACID BINDING.

RK MEDLINE=95179521; PubMed=753044;

RA Klein S., Pelt A., Schner R., Filbin M.T., Tang S., de Bellard M.E., Schner R.L., Mahoney J.A., Hartnell A., Bradfield P., Crocker P.R.;

RT "Sialoadhesin, myelin-associated glycoprotein and CD22 define a new family of sialic acid-dependent adhesion molecules of the immunoglobulin superfamily.";

RL Curr. Biol. 4:965-972(1994).

RL [5]

RP INTERACTION WITH GRB2, SYK, PIK3R1/PIK3R2 AND PLCG1, PHOSPHORYLATION OF TYR-777, TYR-822, TYR-837 AND TYR-857, AND MUTAGENESIS OF TYR-822.

RK MEDLINE=99303650; PubMed=10373493;

RA Yohanan J., Wiens J., Coggeshall K.M., Justement L.B.;

RT "Analysis of tyrosine phosphorylation-dependent interactions between stimulatory effector proteins and the B cell co-receptor CD22.";

RL J. Biol. Chem. 274:18769-18776(1999).

RL [6]

RP INTERACTION WITH GRB2, SHC1 AND INPP5D, AND PHOSPHORYLATION BY LYN.

RK MEDLINE=20298808; PubMed=10748054;

RA Poe J.C., Fujimoto M., Jansen P.J., Miller A.S., Tedder T.F.;

RT "CD22 forms a quaternary complex with SHP, Grb2, and Shc. A pathway for regulation of B lymphocyte antigen receptor-induced calcium flux.";

RL J. Biol. Chem. 275:17420-17427(2000).

RL [7]

RP FUNCTION: Mediates B-cell B-cell interactions. May be involved in the localization of B-cells in lymphoid tissues. Binds sialylated glycoproteins; one of which is CD45. Preferentially binds to alpha2,6-linked sialic acid. The sialic acid recognition site can be masked by cis interactions with sialic acids on the same cell surface. Upon ligand induced tyrosine phosphorylation in the immune response seems to be involved in regulation of B cell antigen receptor signaling. Plays a role in positive regulation

Query Match	Similarity	Score	DB 1	Length	862
Best Local	20.8%	Pred. No. 15			
Matches	55	Conservative 47	Mismatches 105	Indels 58	Gaps 12
32	KKGDTVLLTC--TASQKKSIOFHKNNSNQIKILGNQSSFLTKGSKLNDRAADRSLMDQ	89			
Db	535	RAGGRVLLQCDFAENPAEVRFFPKKNGSLVQEGRYLVSFSVSPE-----DS	581		
Qy	90	GNPFLLIKLNKLIKEDSDTYICEVEDQKEVOLLV-----FGLTANDPHTLLOGQSULTLES	145		
Db	582	GNVYCMWN-----SIGETLSQAMNLQVLLAIPRLRLVRSISGCDVHVECKATLSCS	633		
Qy	146	---PQSSPSVQCKSPRGKNI--QGGKTLVSQVLELQSDGTWTCTVLQNOKKVFEKDIVV	201		
Db	634	DANPFIQIYTWFDSS--GQPLHSGGQKLRLEPLEVQHTGTSVRCKGTGIGTGE-----	684		
Qy	202	LAFQKASSIYKKE--GEQVEFSFLPAFTVEKLTGSGELV-----WQAEKA-----SSS	248		

Db 685 -SPSPSTLVVYSPETIGKRVALGIFLCTICILA----IWGMKIQKKMKONRSQOQHQN 739

QY 249 KSWITPDLKKEYSVKRVTDPLKQ 273

Db 740 SSGQSFVRNKKARPTPLSGPQSQ 764

RESULT 91

NRCA\_CHICK STANDARD; PRT; 1284 AA.

AC P35331;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Neuronal cell adhesion molecule precursor (Nr-CAM) (NGCAM-related cell adhesion molecule) (Ng-CAM-related) (hbravo).

OS Gallus gallus (Chicken).

OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

OC NCBI\_TaxID=9031;

OK [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-52; 178-184 AND 581-594.

RC STRAIN=White leghorn; TISSUE=Embryonic brain;

RX MEDLINE=91258407; PubMed=2045418;

RA Grunet M., Mauro V., Burgeon M.P., Edelman G.M., Cunningham B.A.; "Structure of a new nervous system glycoprotein, Nr-CAM, and its relationship to subgroups of neural cell adhesion molecules.";

RT J. Cell Biol. 113:1399-1412(1991).

RL [2]

RP SEQUENCE OF 25-1284 FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Embryonic brain, and Retina;

RX MEDLINE=92381110; PubMed=1512296;

RA Kayem J.F., Roman J.M., de la Rosa E.J., Schwarz U., Dreyer W.J.; "Bravo/Nr-CAM is closely related to the cell adhesion molecules Ll and Ng-CAM and has a similar heterodimer structure.";

RT J. Cell Biol. 118:1259-1270(1992).

RL -FUNCTION: This protein is a cell adhesion molecule involved in neuron-neuron adhesion, neurite fasciculation, outgrowth of neurites, etc. Specifically involved in the development of optic fibres in the retina.

CC -SUBUNIT: Heterodimer of an alpha and a beta chain.

CC -SUBCELLULAR LOCATION: Type I membrane protein.

CC -ALTERNATIVE PRODUCTS: Event-Alternative splicing; Named isoforms=5; Comment=Additional isoforms seem to exist;

CC Name=1;

CC IsoId=P35331-1; Sequence=Displayed;

CC Name=2; Synonyms=AS10;

CC IsoId=P35331-2; Sequence=VSP\_002603;

CC Name=3; Synonyms=AS12;

CC IsoId=P35331-3; Sequence=VSP\_002604;

CC Name=4; Synonyms=AS93;

CC IsoId=P35331-4; Sequence=VSP\_002605;

CC Name=5; Synonyms=AS-CYT2;

CC IsoId=P35331-5; Sequence=VSP\_002606;

CC -TISSUE SPECIFICITY: Retina and developing brain.

CC -DEVELOPMENTAL STAGE: Expressed in developing neural retina and embryonic brain tissue.

CC -SIMILARITY: Belongs to the immunoglobulin superfamily.

CC -SIMILARITY: Contains 5 fibronectin type III domains.

CC -SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC The European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; X58482; CAA41391.1; -

DR EMBL; L08960; AAA48632.1; -

DR HSPSP: P20241; 1CFB.

DR InterPro; IPR008957; FN\_III-like.

DR InterPro; IPR003961; FN\_III.

DR InterPro; IPR007110; IG\_1like.

DR InterPro; IPR003598; IG\_c2.

DR Pfam; PF00047; fn3; 5.

DR Pfam; PF00047; fn3; 6.

DR SMART; SM00060; FN3; 3.

DR SMART; SM00408; IGc2; 3.

DR PROSITE; PS50835; IG\_Like; 6.

DR KX Immunoglobulin domain; Glycoprotein; signal; Cell adhesion; Repeat; Transmembrane; Alternative splicing.

FT SGNL 1 24

FT CHAIN 25 1284

FT DOMAIN 25 1143

FT TRANSMEM 1144 1166

FT DOMAIN 1167 1284

FT DOMAIN 41 129

FT DOMAIN 136 230

FT DOMAIN 243 332

FT DOMAIN 337 424

FT DOMAIN 430 517

FT DOMAIN 521 608

FT DOMAIN 638 699

FT DOMAIN 738 799

FT DOMAIN 837 906

FT DOMAIN 943 1006

FT DOMAIN 1057 1114

FT DISULFID 63 118

FT DISULFID 162 213

FT DISULFID 268 316

FT DISULFID 358 408

FT DISULFID 452 501

FT DISULFID 543 592

FT CARBOHYD 78 78

FT CARBOHYD 218 218

FT CARBOHYD 290 290

FT CARBOHYD 409 409

FT CARBOHYD 483 483

FT CARBOHYD 576 576

FT CARBOHYD 581 581

FT CARBOHYD 595 595

FT CARBOHYD 692 692

FT CARBOHYD 778 778

FT CARBOHYD 834 834

FT CARBOHYD 885 885

FT CARBOHYD 969 969

FT CARBOHYD 985 985

FT CARBOHYD 995 995

FT CARBOHYD 1048 1048

FT CARBOHYD 1059 1059

FT CARBOHYD 1091 1091

FT CARBOHYD 612 621

FT VARSPLIC 1027 1038

FT VARSPLIC 1039 1131

FT VARSPLIC 1202 1205

FT CONFLICT 209 209

FT CONFLICT 680 680

FT SEQUENCE 1284 AA; 141851 MW; 141851 MW; A570BF9C3D47A0F CRC64;

Query Match 4.5%; Score 104; DB 1; Length 1284;

Best Local Similarity 20.0%; Pred. No. 25;

Matches 91; Conservative 68; Mismatches 172; Indels 124; Gaps 23;

QY 18 LIPATQGNKRVVIGKGGDTVELTCTASQKSIQPHMKNSQIKILNGQSFLLTKGPSKLN 77

Db 247 LTPMGSTNRYEL--RGNVILLLECIAAGLPTVPIRW-----IXEG 284

QY	76	DRADSRSLMDQGNFPLIILKNLKIEDSDP--YICEVEDQKEVQLVGLTANSDFLLQ	135		
Db	285	GELPANKTFPE--NFKKTLKIDVSEADSGNYKCTARN-----TLGSTHVI-	329		
QY	136	QGSILTL-----TLESPGSSPSVQCSPPKGNKIQGGTLVSQLELDQSGTWC	184		
Db	330	--SVTVAAAPYWTITADPNILVLSPGEDGTLICRA-----NGNPKSISWL--TNGVPIA	378		
QY	185	TVLQNKQKVEFKID--IYVLAIFQKASSIYVKKEGQVDFSPFLAFYTK-----LT	233		
Db	379	IAPDPBP--KVDGDTILFSAQGRSSAVYQGNASN--EYGLLANAFVNLAPPRLT	434		
QY	234	GSGLMWAQARASSSSKSWITFDLKNKNEVSKKVTDPKLQMGKLLPLH--LTLPOLP	289		
Db	435	PANKLVGVINDSPALIDCAVFGSPKREIMFPGVGSILR--GNEVHFHDNCTLEIPVA--	491		
QY	290	QYAGSGNLTALAEAKTKLHQEVNVLVMEATOL-----QKULTCEV--KQPTS	335		
Db	492	QKSTGTGYTVVANKNKLGKQNEVQLEVKDPTMIKQPOKYVIGRSAGASPEFCVKHDP	551		
QY	336	PKMLSLKLENKEAKVSKR----EKPVWVLN--DEAGWMOCLSDSQVLESINIKLP	388		
Db	552	IPYIVYWK--DNNEIPDDERFLVGKNDLITMANVTDKDDGYTICVITVITDSVASAV----	606		
QY	389	TWSTPVHPRASALPAPPTGSALPDPQTASALPDP	423		
Db	607	-----LTVVAAPPT-----PALIYARBNP	626		
RESULT 92					
ID	184	VG2_COTVA	STANDARD; PRT; 1348 AA.		
AC	P52583;				
DT	01-OCT-1996	(Rel. 34, Last created)			
DT	01-OCT-1996	(Rel. 34, Last sequence update)			
DT	10-OCT-2003	(Rel. 42, Last annotation update)			
DE	Vascular endothelial growth factor receptor 2 precursor (EC 2.7.1.112)				
DE	(VEGFR-2) (Endothelial kinase receptor EKI) (Qnek 1) (Qnek1).				
GN	KDR OR FLK-1 OR EKI.				
OS	Ccetunrix cctunrix japonica (Japanese quail).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
OC	Coccyzinae.				
OX	NCBI_TaxId=93934;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RP	TISSUE=Embryo;				
RX	MEDLINE=97017121; PubMed=8863722;				
RA	Eichmann A., Marcelle C., Breant C., Le Douarin N.M.;				
RT	"Molecular cloning of Qnek 1 and 2, two quail vascular endothelial				
RL	growth factor (VEGF) receptor-like molecules.";				
RL	Gene 174:3-8(1996).				
RN	[2]				
RP	SEQUENCE OF 910-1348 FROM N.A.				
RP	TISSUE=Spinal cord;				
RX	MEDLINE=93378866; PubMed=8396413;				
RA	Eichmann A., Marcelle C., Breant C., Le Douarin N.M.;				
RT	"Two molecules related to the VEGF receptor are expressed in early				
RT	endothelial cells during avian embryonic development.";				
RL	Mech. Dev. 42:33-48(1993).				
RN	[3]				
RP	SEQUENCE OF 764-880 FROM N.A., AND CHARACTERIZATION.				
RP	TISSUE=Embryo;				
RX	MEDLINE=95301109; PubMed=7781909;				
RA	Flamme I., Breier G., Risau W.;				
RT	"Vascular endothelial growth factor (VEGF) and VEGF receptor 2 (flk-1)				
RT	are expressed during vasculogenesis and vascular differentiation in				
RL	the quail embryo.";				
RL	Dev. Biol. 169:699-712(1995).				
CC	-I- FUNCTION: RECEPTOR FOR VEGF OR VEGF-C. HAS A TYROSINE-PROTEIN				
CC	KINASE ACTIVITY. THE VEGF-KINASE LIGAND/RECEPTOR SIGNALING SYSTEM				
CC	PLAYS A KEY ROLE IN VASCULAR DEVELOPMENT AND REGULATION OF				
CC	VASCULAR PERMEABILITY.				

CC	-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-1- TISSUE SPECIFICITY: IN ALL ENDOTHELIAL TISSUES DURING ONSET OF VASCULARIZATION. IN LATER DEVELOPMENTAL, PRESENT IN LUNG, HEART,
CC	INTESTINE AND SKIN.
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED IN WHOLE MESODERM AT ONSET OF GASTRULATION. FROM DAY 2, CONFINED TO ENDOTHELIAL TISSUES AND
CC	EXPRESSION CONTINUES TO BE WIDESPREAD THROUGHOUT VASCULARIZATION UNTIL E9 WHERE IT BECOMES RESTRICTED TO SPECIFIC REGIONS SUCH AS THE SPINAL CHORD AND HEART VALVES.
CC	-1- INDUCTION: IN VITRO, VEGR IS INDUCED BY BASIC FIBROBLAST GROWTH FACTOR (FGF), UNIQUELY IN THE FIRST 24 H OF CELL CULTURE.
CC	-1- SIMILARITY: Belongs to the Tyr family of protein kinases. csp-1/pdgfr receptor subfamily.
CC	-1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; X83288; CAA58268.1; -;
DR	EMBL; S65205; AAB28127.1; -;
DR	EMBL; S78345; AAB34594.1; -;
DR	PIR; J04953; S51656.
DR	HSSD; P1362; IRGK.
DR	InterPro; IPRO07110; Ig-like.
DR	InterPro; IPRO03598; Ig_c2.
DR	InterPro; IPRO00719; Prot_kinase.
DR	InterPro; IPRO01824; Recepttykiiii.
DR	InterPro; IPRO01245; Tyr_kinase.
DR	InterPro; IPRO08266; Tyr_kinase_AS.
DR	Pfam; PF00047; Ig_6.
DR	Pfam; PF00069; pkinase; 1.
DR	Prodrom; PD000001; Prot_kinase; 2.
DR	SMART; SMO0408; IGC2; 1.
DR	SMART; SMO0219; Tyrc; 1.
DR	PROSITE; PS50835; IG_LIKE; 5.
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR	PROSITE; PS500109; PROTEIN_KINASE_TYR; 1.
KM	PROSITE; PS00240; RECEPTOR_TYR_KIN_III; 1.
KM	Angiogenesis; Signal; Transferrase; Tyrosine-protein kinase; Receptor;
KW	Transmembrane; Glycoprotein; Phosphorylation; ATP-binding;
KV	Immunoglobulin domain; Repeat.
FT	SIGNAL 1 20
FT	CHAIN 21 1348
FT	POTENTIAL.
FT	VASCULAR ENDOTHELIAL GROWTH FACTOR
FT	RECEPTOR 2.
FT	DOMAIN 21 756
FT	TRANSMEM 757 777
FT	POTENTIAL.
FT	EXTRACELLULAR (POTENTIAL).
FT	CYTOPLASMIC (POTENTIAL).
FT	IG-LIKE C2-TYPE 1.
FT	IG-LIKE C2-TYPE 2.
FT	IG-LIKE C2-TYPE 3.
FT	IG-LIKE C2-TYPE 4.
FT	IG-LIKE C2-TYPE 5.
FT	IG-LIKE C2-TYPE 6.
FT	IG-LIKE C2-TYPE 7.
FT	PROTEIN KINASE.
FT	NP_BIND 831 839
FT	BINDING 859 859
FT	ACT SITE 1021 1021
FT	CARBONYD 43 43
FT	N-LINKED (GLCNAC. . .)
FT	(POTENTIAL).
FT	CARBONYD 47 47
FT	N-LINKED (GLCNAC. . .)
FT	(POTENTIAL).
FT	CARBONYD 63 63
FT	N-LINKED (GLCNAC. . .)
FT	(POTENTIAL).
FT	CARBONYD 93 93
FT	N-LINKED (GLCNAC. . .)
FT	(POTENTIAL).
FT	CARBONYD 138 138
FT	N-LINKED (GLCNAC. . .)
FT	(POTENTIAL).
FT	CARBONYD 153 153
FT	N-LINKED (GLCNAC. . .)
FT	(POTENTIAL).
FT	CARBONYD 201 201
FT	N-LINKED (GLCNAC. . .)
FT	(POTENTIAL).



Query Match	Similarity	Score	DB	Length
Best Local	Similarity	Score	DB	Length
Matches	73	Conservative	55	Mismatches 130; Indels 122; Gaps 19
FT	CARBOHYD	240	240	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	290	290	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	310	310	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	365	365	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	386	386	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	513	513	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	556	556	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	603	603	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	613	613	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	622	622	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	666	666	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	688	688	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CARBOHYD	710	710	N-LINKED (GLCNAC . . .) (POTENTIAL)
FT	CONFLICT	965	865	A -> S (IN REF 3)
SO	SEQUENCE	1348	AA; 150305	MM; A5E4194A76DF5B3 CRC64;
Query Match	Similarity	Score	DB	Length
Best Local	Similarity	Score	DB	Length
Matches	73	Conservative	55	Mismatches 130; Indels 122; Gaps 19
QY	9	HLVLVQLALLPAAATQGNKVVLGKKGDTVELTCTASQKKSIOQFW-KNSNQIKLGNQGS	67	
DB	536	HTGGLGLEINLQPRSQLTEK-----DNTSLQCTADKFTFEKLSWYKLSYHV-----SQTP	584	
QY	68	FLITG--SKUNDRADSRSL-----WDGNPFLIKKLIKIDSDTYICEVDQOE	117	
DB	585	F---GGELMPCKKNLADLOKLNATVSNNGEENVLELLLRNLSLDGDDYVCIAODKAK	641	
QY	118	VQ-LVFLFELTANSDTH-----LLOGQSITLTLESPPGSSPVQCRSPRG-----KN	162	
DB	642	TQHCIVKRLTYOERPLHRLVNGLENOTINI-----GETIYLC-TVNGVPPRNTTWERN	694	
QY	163	IQ-----GKTLVSQLELODSGTWTCTVLONQKKEFKIDIVLAFQKASSI	210	
DB	695	SETLPEDSGIVLKQDNKTKTLTRVAKKEGGLVYTCACN-----ILGCKKABAF	742	
QY	211	VYKKEGEQEVESFPPLATFYEKLTGSGELMW-----QAEASSSKSWITPPLXKE	260	
DB	743	F-SVQGAEEKTNLELLILVGTAVIAMFPLLVIITLRTVKRANGGDMKTGYSTIMDPE	801	
QY	261	VSV-----KRVATOD-----PKLQMGKKLPLHLTPQALPOLYAGSGNLTALBAKTKL	308	
DB	802	VPIDEHCERLPLPDASKMEFPRDLKLGKPL-----GRGAFGVTEAD-----	843	
QY	309	HOENVLVVWRATQLOKKNLTC	328	
DB	844	-----AFGIDXTATC	853	
RESULT 93				
NEOI_HUMAN				
ID	NEOI_HUMAN	STANDARD;	PRT:	1461 AA.
AC	Q92859;	000340;		
DT	16-OCT-2001	(Rel. 40, Created)		
DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	Neogenin precursor.			
GN	NEOI OR NGN.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).			
RC	TISSUE=Fetal brain;			
RX	MEDLINE=97236653; PubMed=9121761;			
RA	Meयरhard J.A., Look A.T., Bigner S.H., Fearon E.R.;			
RT	"Identification and characterization of neogenin, a DCC-related			
RL	gene."			
RP	Oncogene 14:1129-1136 (1997).			
RC	[2]			
RC	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).			
RC	TISSUE=Fetal brain;			

RX	MEDLINE=97312699; PubMed=9169140;
RA	Vielmester J., Chen X.-N., Mikelvich F., Lane R.P., Yamakawa K.,
RB	Korenberg J.R., Dreyer W.U.;
RT	"Molecular characterization of human neogenin, a DCC-related protein,
RT	and the mapping of its gene (NEO1) to chromosomal position 15q22.3-
RT	q23.";
RL	Genomics 41:414-421(1997).
CC	-I- FUNCTION: May be involved as a regulatory protein in the
CC	transition of undifferentiated proliferating cells to their
CC	differentiated state. May also function as a cell adhesion
CC	molecule in a broad spectrum of embryonic and adult tissues.
CC	-I- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-I- ALTERNATIVE PRODUCTS:
CC	Event-Alternative splicing; Named isoforms=2;
CC	Comment=Additional isoforms seem to exist;
CC	Name=1;
CC	IsoId=Q92859-1; Sequence=Displayed;
CC	Name=2;
CC	IsoId=Q92859-2; Sequence=VSP_002593;
CC	-I- TISSUE SPECIFICITY: Widely expressed and also in cancer cell
CC	lines.
CC	-I- SIMILARITY: Belongs to the immunoglobulin superfamily. DCC family.
CC	-I- SIMILARITY: Contains 4 immunoglobulin-like C2-type domains.
CC	-I- SIMILARITY: Contains 6 fibronectin type III domains.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on tests
CC	use by non-profit institutions as long as its content is in no way com-
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a> )
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; U61262; AAB17263.1; .
DR	EMBL; U72391; AAC51287.1; .
DR	HSSP; P02751; ITTF.
DR	GeneW; HGNC:7754; NEO1.
DR	MIM; 601907; .
DR	GO; GO:0005887; C:integral to plasma membrane; TAS.
DR	InterPro; IPRO08957; FN.III-like.
DR	InterPro; IPRO03961; FN.III.
DR	InterPro; IPRO03962; FN.III subd.
DR	InterPro; IPRO07110; Ig-Like.
DR	InterPro; IPRO03598; Ig_C2.
DR	Pfam; PF00047; IG; 4.
DR	Pfam; PF00047; IG; 6.
DR	PRINTS; PR00014; FNTYPEIII.
DR	SMART; SMO0060; FN3; 6
DR	SMART; SMO0408; IGC2; 3.
DR	PROSITE; PS50835; IG_LIKE; 4.
KW	Cell adhesion; Repeat; Signal; Transmembrane; Immunoglobulin domain;
KW	Glycoprotein; Alternative splicing.
FT	SIGNAL 1 33 POTENTIAL.
FT	CHAIN 34 1461 NEOGENIN.
FT	DOMAIN 34 1105 EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 106 1126 POTENTIAL.
FT	DOMAIN 1127 1461 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 52 141 IG-LIKE C2-TYPE 1.
FT	DOMAIN 152 238 IG-LIKE C2-TYPE 2.
FT	DOMAIN 243 336 IG-LIKE C2-TYPE 3.
FT	DOMAIN 341 426 IG-LIKE C2-TYPE 4.
FT	DOMAIN 436 533 FIBRONECTIN TYPE-III 1.
FT	DOMAIN 536 629 FIBRONECTIN TYPE-III 2.
FT	DOMAIN 630 729 FIBRONECTIN TYPE-III 3.
FT	DOMAIN 735 829 FIBRONECTIN TYPE-III 4.
FT	DOMAIN 850 950 FIBRONECTIN TYPE-III 5.
FT	DOMAIN 951 1052 FIBRONECTIN TYPE-III 6.
FT	DOMAIN 1118 1121 POLY-VAL.
FT	DISULFID 74 129 BY SIMILARITY.
FT	DISULFID 173 221 BY SIMILARITY.
FT	DISULFID 270 320 BY SIMILARITY.
FT	DISULFID 362 410 BY SIMILARITY.
FT	CARBOND 73 73 N-LINKED (GLCNAC. .) (POTENTIAL).





```

FT DOMAIN 23 45 FRAMEWORK-1.
FT DOMAIN 46 56 COMPLEMENTARITY-DETERMINING-1.
FT DOMAIN 57 71 FRAMEWORK-2.
FT DOMAIN 72 78 COMPLEMENTARITY-DETERMINING-2.
FT DOMAIN 79 110 FRAMEWORK-3.
FT DOMAIN 111 119 COMPLEMENTARITY-DETERMINING-3.
FT DOMAIN 120 129 FRAMEWORK-4.
FT DISULFID 45 110 BY SIMILARITY.
FT NON_TER 129 129
SQ SEQUENCE 129 AA; 14069 MW; F941FA07D4AFCE29 CRC64;

Query Match
Best Local Similarity 4.5%; Score 103.5; DB 1; Length 129;
Matches 43; Conservative 10; Mismatches 61; Indels 17; Gaps 5;

QY 1 MNRGVPFRLLVQLALPA-----TGNKVVLAKKQDYELTCTASQKSIQPHWK 54
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 1 MDRKVPQAQ--LIGLLMLRGARCDIQMTQSPSLASVGDVYITCRASQSIISNYLWY 58
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 55 NSNQIKILGNQSGFLTGSPKLNDRADRSRLMDQG--NPLIIKNIKIEDSDTYICEVED 113
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 59 QQKRP---GKAPKLLIYAASLSQSGVTSRSGSGSGTFTLTLSLQPEDSATVYC--- 110
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 114 QKEEVQLVPG 124
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 111 QQSSTLTITFG 121
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 97
NTRI MOUSE STANDARD; PRT; 344 AA.
ID NTRI MOUSE
AC 099PTD;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurotrophin precursor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
OX
(1)
RP SEQUENCE FROM N.A.
RC STRAIN=ICR; TISSUE=Brain;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heife F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Szelesten M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzyzanski M.I., Skaleks U., Smallus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: Neural cell adhesion molecule.
-i- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.

```

```

CC -!- SIMILARITY: Belongs to the immunoglobulin superfamily. IgON
family.
CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF282980; AAK00276.1; -.
CC EMBL; BC023307; AAH23307.1; -.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003599; Ig.
CC InterPro; IPR003598; Ig_c2.
CC Pfam; PF00047; Ig_3.
CC SMART; SM00409; Ig_3.
CC SMART; SM00408; IGC2; 3.
CC PROSITE; PS50835; IG_LIKE; 3.
CC Immunoglobulin domain; Cell adhesion, Glycoprotein; GPI-anchor;
CC Repeat; Signal; Lipoprotein.
CC CHAIN 1 31
CC SIGNAL 1 31
CC PROPER 32 344
CC DOMAIN 39 126
CC DOMAIN 136 218
CC DOMAIN 222 309
CC DISULFID 57 115
CC DISULFID 157 201
CC DISULFID 243 295
CC CARBOHYD 44 70
CC CARBOHYD 70 152
CC CARBOHYD 152 284
CC CARBOHYD 284 292
CC CARBOHYD 305 321
CC CARBOHYD 321 321
CC LIPID 321 321
CC CONFLICT 75 75 L -> P (IN REF. 1).
CC CONFLICT 92 92 S -> G (IN REF. 1).
CC CONFLICT 119 119 T -> I (IN REF. 1).
CC CONFLICT 187 187 E -> Q (IN REF. 1).
CC CONFLICT 213 213 R -> P (IN REF. 1).
CC CONFLICT 225 225 I -> F (IN REF. 1).
CC SEQUENCE 344 AA; 37984 MW; C885BBA52C148554 CRC64;

Query Match
Best Local Similarity 4.5%; Score 103.5; DB 1; Length 344;
Matches 75; Conservative 55; Mismatches 166; Indels 73; Gaps 13;

QY 11 LVLQALPAATQGNKVVIG-----KKGDVELTCTASQKSIQPHWKN 57
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 14 LVVLSRLFLVPTGVPAVSGDATFPKAMDVTVRGEGSATLRCTIDNRTV-RVAVLNRS 72
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 58 QIKILGNQSGFLTGSPKLNDRADRSRLMDQGNFPLIIKNIKIEDSDTYICEVED 117
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 73 TLVANDKWCIDPVRVLN-----TQVYSIEIQNDVDYDEGPTCSVQTDNHP 123
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 118 -----VOLLVFGLTNSDTHLQGSITLTLESPSSPSVQCR--SPRKN-IGGG 166
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 124 KTSRVLIVQVSPKIVYEISDSINEGNNISLTCTAGPEPVTYKRAHSFPAKVGVSVD 183
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 167 KTLVSQELQDSGTCTVQLQNKVEFKIDIVLAFQKASSIVYKGEQVEFSPPLA 226
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 184 EYELIGIRREGQGEVGEASND-----VVAAPVAR--VKVTAVYPPY 224
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 227 FTYEKLITG-----SGELMQAEPASSK-SWTFDLKAKVEVYKVTQDPKLOMKKLP 280
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DQ 225 ISHAKGTVPVGQKTLQCEASAVPSAEFQWPKD-----KRLVEGKGVKVENRPF 276
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```

QY 281 HLTPQLPYAGSGNTLTLAETKGLHGVNLVWRATOLQKNTLCE-----VWGPTSP 336
DB 277 LSKLTFPNVSEHDYGNITCVASNLGHTNASIMLFGGAVSEVNNGTSRAGCIW--LLP 334
QY 337 KLMISLKE 345
DB 335 LVLVHLHLK 343

RESULT 98
NTRI_RAT
ID NTRI_RAT STANDARD; PRT; 344 AA.
AC 062718;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurotrophin precursor (GP65).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 217-229.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=95198094; PubMed=7891157;
RA Struyk A.F., Canoll P.D., Wolfgang M.J., Rosen C.L., D'Eustachio P.,
RA Salzer J.L.;
RT "Cloning of neurotrophin defines a new subfamily of differentially
RT expressed neural cell adhesion molecules.";
RL J. Neurosci. 15:2141-2156(1995).
CC -1- FUNCTION: Neural cell adhesion molecule.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- TISSUE SPECIFICITY: Central nervous system.
CC -1- DEVELOPMENTAL STAGE: Expressed at high levels in several
CC subplate, and lower cortical laminae in the forebrain and in the
CC pontine nucleus, cerebellar granule cells, and Purkinje cells in
CC the hindbrain.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. IGLON
CC family.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U16845; AAA67445.1; -.
DR PIR; I56551; I56551.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003598; IG_c2.
DR Pfam; PF00047; IG; 3.
DR SMART; SM00408; IGc2; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
KW Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW Repeat; Signal; L1poprotein.
KM
FT CHAIN 1 321
FT SIGNAL 31
FT PROPEP 322 344
FT DOMAIN 39 126
FT DOMAIN 136 218
FT DOMAIN 222 309
FT DISULFID 57 115
FT DISULFID 157 201
FT CARBOHYD 243 295
FT CARBOHYD 44 44
FT CARBOHYD 70 70
FT CARBOHYD 152 152
FT CARBOHYD 216 216
FT CARBOHYD 284 284
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).
N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 292 292
FT CARBOHYD 305 305
FT CARBOHYD 321 321
FT LIPID 321 321
SQ SEQUENCE 344 AA; 37998 MW; CBB39BE53B3B224 CRC64;

Query Match
Best Local Similarity 20.1%; Pred. No. 5;
Matches 74; Conservative 54; Mismatches 168; Indels 73; Gaps 13;

QY 11 LVLVQLALPLPAATQGNKRVVG-----KKGDVVELTCTASQKKSIOFHMKN 57
DB 14 LVLVSLRLFLVPTGVVRSGDATFPKAMDNVYVQGESATLCTIDNRT-RAVLMNR 72
QY 58 QIKILNQGSFLLTKGSKLNDRADSRRLMDQGNFPLIKLKIEDSDTYICEVDQKE 117
DB 73 TLVYAGDKMCLDPRVLLSN-----TQTOYSIEIQNVVDYEGPYTCGVTDNHP 123
QY 118 -----VOLLVFGLTANSDFHLQGSLLTLIESPSSPSVQCR--SPRKN-IGG 166
DB 124 KTSRVHLIVVSPKIVEISSDISINEGNISLTCTIATGRPEPTVWRHISPKAVGVS 183
QY 167 KTLVSQLELDSDGTWCTVLQKQKVEFKIDIVLAFKASSIVYKKEGOVEFSPPLA 226
DB 184 EYLEIQITREQSGEYVCSASN-----DVAAPVVRVN-----VTNVPY 224
QY 227 FVYEKLTG-----SGEIMWAERASSK-SWTFDLKNKEVSKYKTDPLQMGKPL 280
DB 225 ISEAKGVGVVQKGLQCEASAVPSAEFQFQD-----KRLVEGKGVAVENRPF 276
QY 281 HLTPQLPYAGSGNTLTLAETKGLHGVNLVWRATOLQKNTLCE-----VWGPTSP 336
DB 277 LSKLTFPNVSEHDYGNITCVASNLGHTNASIMLFGGAVSEVNNGTSRAGCIW--LLP 334
QY 337 KLMISLKE 345
DB 335 LVLVHLHLK 343

RESULT 99
GAG_SIVAT
ID GAG_SIVAT STANDARD; PRT; 519 AA.
AC POS889;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE GAG polypeptide [Contains: Core proteins P17, P24, and P15].
GN GAG.
OS Simian immunodeficiency virus (TYO-1 isolate) (SIV-AGM).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11731;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88232906; PubMed=3374586;
RA Pukaew M., Mura T., Hasegawa A., Morikawa S., Tsujimoto H.,
RA Miki K., Kitamura T., Hayami M.;
RT "Sequence of simian immunodeficiency virus from African green monkey,
RT a new member of the HIV/SIV group.";
RL Nature 333:457-461(1988).
CC -1- MISCELLANEOUS: This is an African green monkey isolate.
CC -1- SIMILARITY: Contains 2 CCHC-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X07805; CAA30657.1; -.
DR HSP; P05888; 1AA.

```

```

DR HTV; X07805; GAGSAGMVT.
DR InterPro: IPR000721; Gag_p24.
DR InterPro: IPR008916; Retrov_capsid_C.
DR InterPro: IPR008919; Retrov_capsid_N.
DR InterPro: IPR000071; Retrov_p17.
DR InterPro: IPR001878; Znf_CCHC.
DR Pfam: PF00540; Gag_p17; 1.
DR Pfam: PF00607; Gag_p24; 1.
DR Pfam: PF00098; Zf-CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PRINTS; PR00234; HIVMTRIX.
DR SMART; SM00343; Znf_CCHC; 2.
DR PROSITE; PS50158; Zf_CCHC; 2.
KM AIDS; Core protein; Polypeptide; Zinc-finger; Repeat.
FT CHAIN 1 141 CORE PROTEIN P17.
FT FTN 142 519 CORE PROTEIN P17.
FT ZN_FING 397 414 CCHC-TYPE 1.
FT ZN_FING 418 435 CCHC-TYPE 2.
SQ SEQUENCE 519 AA; 58143 MW; 85A3AC06BCCDCA38 CRC64;

Query Match 4.5%; Score 103.5; DB 1; Length 519;
Best Local Similarity 19.6%; Pred. No. 8.5;
Matches 101; Conservative 62; Mismatches 162; Indels 191; Gaps 26;

QY 18 LLEPAQGNVVLG-----KKGDTVELTCAQCKSIQPHMKSNQIKILGNQ 65
DB 67 LEPTSGEGLSLNLYCVLYCLHKEQKVYKTEEAVALRQCHLVEKESATE----- 119
QY 66 GSEFLTKGPKSLNDR-----ADSRSLMDQGNFLLIKLIKIEDSDTYICEVE 112
DB 120 -----TSSGQKKNKXGIAAPGCGQNFPAQOQGNAM--VHPLPRLT-----NAMKAVE 168
QY 113 DQK-BEV-----QLVFGLTANSDTLL-----QCGSLTLTLESP 147
DB 169 EKFKGAELIVMFAISGCTPPYDINQMLNVLGDHOGALQVKEIINEEAAQMWVTHPLPA 228
QY 148 GSSPSVOCSPRGNIQGGKTLVSQLELQDSGTWTCTVLQNKQKVEFKIDI-----V 200
DB 229 GRLPAQGLRDPKRSIDIGTTSSVOEQL-----WITT-----ANPRVDGALIRRWI 275
QY 201 VLAFOKA-----SSIVYKKEGQVEPS-----FPLAFTEKLTGSGELMWQAEKASSSK 249
DB 276 ILGLQCKVKMYNPVSLDINGQREPRKQVDRFYKAIKRAEQ--ASGEV-----K 323
QY 250 SATTFDLAKKEVSKVTOPKQLOMGKLLPLHLTLPALQVYAGSGNLTALAKTKGLH 309
DB 324 QWMT-----ESLTIQNNAPDCKVYL-KGLQMHPTLEMLTACQGVG-----GPSY 367
QY 310 QEVNLYVVRATQLOKNTCEWVGPTSPKMLSLKLEKKAQVSKREK-PVAVLN----- 362
DB 368 KAKVMAEMQTMQNNQNV-QQGGP-----KQGRPLRCTNCGKFGH 407
QY 363 -----PEAGMOCCLSDS-----GOV-----LLESNIKVLPTWSTPVHPRASAL 401
DB 408 MOROCPEPRKTKCKKCGKGLHAKDCQGVNPLGYGRWMAKPRNFPATLGAEPSSA--- 464
QY 402 PAPTGSALPDP-----QTSALPDP 424
DB 465 PPPSGTTPYDPAKQLQOYAEKQKQLREQKRNP 500

RESULT 100
PCL_HUMAN STANDARD; PRT; 738 AA.
AC P16284;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Platelet endothelial cell adhesion molecule precursor (PECAM-1)
DE (CD31 antigen) (EndoCAM) (GPIIb).
GN PECAM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90193682; PubMed=1690453;
RA Newman P.J., Berndt M.C., Gorstki J., White J.C. II, Lyman S.,
RA Padock C., Muller W.A.;
RT "PECAM-1 (CD31) cloning and relation to adhesion molecules of the
RT immunoglobulin gene superfamily.",
RN Science 247:1219-1222 (1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91340830; PubMed=1874786;
RA Albelda S.M., Muller W.A., Buck C.A., Newman P.J.;
RT "Molecular and cellular properties of PECAM-1 (endocam/CD31): a novel
RT vascular cell-cell adhesion molecule.",
RN J. Cell Biol. 114:1059-1068 (1991).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91060975; PubMed=1700999;
RA Stockinger H., Gadd S.J., Eher R., Majdic O., Kasinrek W.,
RA Schneider W., Strass B., Schnabl E., Knapp W.;
RT "Molecular characterization and functional analysis of the leukocyte
RT surface protein CD31.",
RN J. Immunol. 145:3889-3897 (1990).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=95086194; PubMed=7994021;
RA Kriesbaum N.E., Gumina R.J., Newman P.J.;
RT "Organization of the gene for human platelet/endothelial cell
RT adhesion molecule-1 shows alternatively spliced isoforms and a
RT functionally complex cytoplasmic domain.",
RN Blood 84:4028-4037 (1994).
RN [5]
RP VARIANT VAL-125.
RX MEDLINE=96133740; PubMed=8532023;
RA Behar E., Chao N.J., Hiraki D.D., Krishnaswamy S., Brown B.W.,
RA Zehnder J.L., Grunet F.C.;
RT "Polymorphism of adhesion molecule CD31 and its role in acute
RT graft-versus-host disease.",
RN New Engl. J. Med. 334:286-291 (1996).
RN [6]
RP VARIANT VAL-125.
RX MEDLINE=21649465; PubMed=11791967;
RA Casals-Pascual C., Allen S., Allen A., Kai O., Lowe B., Pain A.,
RA Roberts D.V.;
RT "Codon 125 polymorphism of CD31 and susceptibility to malaria.",
RN Am. J. Trop. Med. Hyg. 65:736-737 (2001).
CC -1- FUNCTION: THIS PROTEIN IS A CELL ADHESION MOLECULE EXPRESSED ON
CC PLATELETS AND AT ENDOTHELIAL CELL INTERCELLULAR JUNCTIONS.
CC -1- SUBCELLULAR LOCATION: TYPE I membrane protein.
CC -1- PFM: PHOSPHORYLATED ON SER AND TYR RESIDUES AFTER CELLULAR
CC ACTIVATION.
CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD31 entry;
CC WWW=http://www.ncbi.nlm.nih.gov/prov/cd/cd31.htm".
CC -1- DATABASE: NAME=RED Systems' cytokine source book: PECAM1;
CC WWW="http://www.rndsystems.com/asp/g_sitebuilder.asp?bodyd=218".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M28526; AAA36429.1; -
DR EMBL; M37780; AAA36186.1; -
DR EMBL; L34657; AAA60057.1; -
DR EMBL; L34631; AAA60057.1; JOINED.
DR EMBL; L34637; AAA60057.1; JOINED.
DR EMBL; L34638; AAA60057.1; JOINED.

```

[illegible]

Db 466 DNPTEDEVEXQ 475

Search completed: August 3, 2004, 13:09:14  
Job time : 14.221 secs

	Query March	4.5%	Score 103.5;	DB 1;	Length 738;
	Best Local Similarity	25.8%;	Pred. No. 13;		
	Matches	49;	Conservative	30;	Mismatches    64; Indels    27; Gaps    8.
Oy	101 IEDSDTYICEVEDOK-EEVOLIFELT-----ANSDTHLLQCSSTLTLESPGSSP	151			
Dd	296 VEHSGNVCYKVESSISIKVSSIVVNITELFKPELLESSFTHHDDGERLNLSG-SIRGNRP	354			
Oy	152 SVQCNSPRGKNIOGSKTLVSOLF-----LDSCGTWCTCY-LONOKKVYEKKIDIVLA	203			
Dd	355 A-----NPTIQKEDI-VSOTDOPFTRIAKSMSDGITYICAGIDIKVVKSNVTQIWC	406			
Oy	204 FOKASIIYVKKEGEVERFSPLAFIVEKLTGSGELMMOAEKRASSKSWTTDPDKKEYSV	263			
Dd	407 MLSQRISIDPAOFEVIK-QGITLEVCEGISGLPIPSYLKLTISKVLLENSTAKNSNDPAVFK	465			
Oy	264 KRVTDPPKLO	273			

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:51:44 ; Search time 31.1057 Seconds

(without alignments)  
4594.975 Million cell updates/sec

Title: SEQ5  
Perfect score: 2325  
Sequence: 1 MNRGVPRHLLVQLALP.....VISFLGLGVACVLRTR 453

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 125 summaries

Database :

1: SPREMBL.25.\*  
2: sp\_archaea.\*  
3: sp\_bacteria.\*  
4: sp\_fungi.\*  
5: sp\_human.\*  
6: sp\_invertebrate.\*  
7: sp\_mammal.\*  
8: sp\_mhc.\*  
9: sp\_organelle.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriopl.\*  
17: sp\_archaeopl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1715	73.8	397	6	009261
2	1713	73.7	397	6	009260
3	1711	73.6	397	6	095N99
4	1708	73.5	397	6	009259
5	1706	73.4	397	6	009263
6	1699	73.1	397	6	009262
7	1699	73.1	397	6	002805
8	1615.5	69.5	457	6	08H278
9	1585.5	68.2	457	6	08H277
10	1188	51.1	455	6	09X578
11	1087	46.8	474	6	P79355
12	985	42.4	457	11	Q61396
13	954.5	41.1	433	11	055054
14	400	17.2	86	6	077596
15	400	17.2	86	6	077597
16	397	17.1	86	6	077594

17	397	17.1	86	6	077599	077599 theropithec
18	395	17.0	86	6	077595	077595 cercopithec
19	392	16.9	86	6	077598	077598 papio sp. (
20	388	16.7	86	6	077601	077601 lophocobus
21	383	16.5	86	6	077600	077600 lophocobus
22	357	15.4	71	4	Q13969	Q13969 homo sapien
23	305.5	13.1	99	6	Q29027	Q29027 sus scrofa
24	280.5	12.1	99	6	Q29028	Q29028 sus scrofa
25	278.5	12.0	120	11	Q88650	Q88650 marmota mon
26	271.5	11.7	482	13	Q90MB5	Q90MB5 anas platyr
27	245	10.5	487	13	Q9M6V7	Q9M6V7 gallus gall
28	237	10.2	90	11	P70443	P70443 mus musculu
29	174	7.5	739	6	Q28260	Q28260 canis famli
30	165	7.1	650	6	Q9GKR2	Q9GKR2 bos taurus
31	165	7.1	739	6	Q9GKR3	Q9GKR3 bos taurus
32	162	7.0	739	11	Q63669	Q63669 rattus norv
33	161.5	6.9	538	6	Q29123	Q29123 sus scrofa
34	157	6.8	955	4	Q8NPP4	Q8NPP4 homo sapien
35	155.5	6.7	538	6	Q28939	Q28939 sus scrofa
36	152	6.5	31	4	Q9UDE5	Q9UDE5 homo sapien
37	152	6.5	1496	4	Q92626	Q92626 homo sapien
38	151.5	6.5	739	6	Q865F2	Q865F2 oryctolagus
39	145	6.2	1431	11	Q80U60	Q80U60 mus musculu
40	144.5	6.2	437	4	Q81ZP8	Q81ZP8 homo sapien
41	144	6.2	760	11	Q91X98	Q91X98 mus musculu
42	143	6.2	5636	4	Q96RW7	Q96RW7 homo sapien
43	142	6.1	739	11	Q8K0X1	Q8K0X1 mus musculu
44	142	6.1	885	6	Q8H7V1	Q8H7V1 sus scrofa
45	142	6.1	886	6	Q8HYV2	Q8HYV2 sus scrofa
46	141.5	6.1	397	11	Q9JIE0	Q9JIE0 mus musculu
47	140.5	6.0	702	4	Q8N4D0	Q8N4D0 homo sapien
48	140.5	6.0	17352	5	Q95YM2	Q95YM2 procamburus
49	139.5	5.9	403	5	Q9VP08	Q9VP08 drosophila
50	138	5.9	398	11	Q07763	Q07763 mus musculu
51	138	5.9	2828	4	Q9NR99	Q9NR99 homo sapien
52	137.5	5.9	1259	11	Q9QY38	Q9QY38 mus musculu
53	136.5	5.9	1395	5	Q44924	Q44924 drosophila
54	136.5	5.9	1465	11	Q770G5	Q770G5 mus musculu
55	136.5	5.9	4117	5	Q8IRV9	Q8IRV9 drosophila
56	136.5	5.9	4179	5	Q9W4Y4	Q9W4Y4 drosophila
57	136.5	5.9	4223	5	Q8MPN3	Q8MPN3 drosophila
58	136.5	5.9	4228	5	Q8IRV8	Q8IRV8 drosophila
59	136	5.8	512	4	Q96DN8	Q96DN8 homo sapien
60	136	5.8	545	5	Q9VCT4	Q9VCT4 drosophila
61	134.5	5.8	1035	5	Q9NEG1	Q9NEG1 drosophila
62	134	5.8	487	13	Q772H2	Q772H2 gallus gall
63	134	5.8	1445	11	Q63155	Q63155 rattus norv
64	133.5	5.7	6620	4	Q96AA2	Q96AA2 homo sapien
65	133	5.7	2693	5	Q8ISF3	Q8ISF3 cercopithec
66	133	5.7	2708	5	Q8ISF4	Q8ISF4 cercopithec
67	133	5.7	2780	5	Q8MNS0	Q8MNS0 caenorhabdi
68	133	5.7	2808	5	Q8MNS1	Q8MNS1 caenorhabdi
69	133	5.7	18519	5	Q8ISF6	Q8ISF6 caenorhabdi
70	133	5.7	18534	5	Q8ISF7	Q8ISF7 caenorhabdi
71	132.5	5.7	352	13	Q9W6V2	Q9W6V2 gallus gall
72	132.5	5.7	1596	4	Q9HCL6	Q9HCL6 homo sapien
73	132	5.7	811	5	Q9VNP2	Q9VNP2 drosophila
74	132	5.7	1011	5	Q24273	Q24273 drosophila
75	132	5.7	1248	6	Q9X741	Q9X741 cercopithec
76	132	5.7	1255	6	Q7YOL7	Q7YOL7 pongo pygma
77	132	5.7	2212	4	Q8NHN3	Q8NHN3 homo sapien
78	132	5.7	6831	5	Q23550	Q23550 caenorhabdi
79	132	5.7	7158	5	Q23551	Q23551 caenorhabdi
80	131.5	5.7	848	5	Q25198	Q25198 hydra atten
81	131	5.6	483	13	Q7SX76	Q7SX76 brachydanio
82	131	5.6	827	11	Q80TU8	Q80TU8 mus musculu
83	131	5.6	1255	4	Q723Z9	Q723Z9 homo sapien
84	131	5.6	1255	4	Q7YOL8	Q7YOL8 pan troglod
85	131	5.6	1340	4	Q8NDA2	Q8NDA2 homo sapien
86	131	5.6	5175	5	Q810L3	Q810L3 caenorhabdi
87	131	5.6	5198	5	Q76518	Q76518 caenorhabdi
88	130	5.6	4162	13	Q98918	Q98918 gallus gall
89	129.5	5.6	340	11	Q88654	Q88654 mus musculu

```

90      129      5.5      287      11      Q9D7B8
91      128.5      5.5      564      13      Q7ZU00
92      128.5      5.5      1395      5      Q9M213
93      128.5      5.5      1413      11      Q8GCB3
94      127      5.5      2629      11      Q8G533
95      126.5      5.4      399      11      Q9SE55
96      126.5      5.4      484      5      Q26475
97      126      5.4      785      11      Q7TNP4
98      126      5.4      3197      5      Q9MID5
99      125.5      5.4      513      4      Q00481
100      125      5.4      605      11      Q921P2
101      125      5.4      838      11      Q8C4B2
102      125      5.4      838      11      Q8BQ96
103      125      5.4      1056      13      Q7ZW34
104      125      5.4      1746      4      Q8WY19
105      125      5.4      6875      6      Q28733
106      124.5      5.4      355      4      Q7Z3B1
107      124.5      5.4      358      13      Q8UV81
108      124.5      5.4      536      11      Q8BJE2
109      124.5      5.4      544      13      Q7Z285
110      124      5.3      520      11      Q925P2
111      124      5.3      542      4      Q8NHN5
112      124      5.3      731      6      Q8SP16
113      124      5.3      758      6      Q9N3H7
114      124      5.3      7962      4      Q10465
115      124      5.3      34350      4      Q8W242
116      123.5      5.3      325      13      Q8UV52
117      123      5.3      18412      13      Q7Z261
118      122.5      5.3      338      4      Q81V49
119      122.5      5.3      495      4      Q9HCY1
120      122.5      5.3      733      6      Q8SQ83
121      122.5      5.3      2013      11      Q9ERC8
122      122      5.2      859      15      Q92937
123      122      5.2      1333      13      Q08476
124      122      5.2      2013      11      Q8VHZ8
125      121.5      5.2      577      11      Q9D221

```

## ALIGNMENTS

```

RESULT 1
009261      PRELIMINARY;      PRT;      397 AA.
AC      009261;
DT      01-JUL-1997 (TrEMBLrel. 04, Created)
DT      01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
GN      CD4 (Fragment).
OS      Cercopithecus sabaenus.
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC      Cercopithecinae; Cercopithecus.
OX      NCB1_TaxID=60711;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=98017879; PubMed=9379478;
RA      Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Machiot C.,
RA      Corbet S., Barre-Sinoussi F., Allan J.S.;
RA      "Relation between phylogeny of African green monkey CD4 genes and
RA      their respective simian immunodeficiency virus genes.";
RL      J. Med. Primatol. 26:120-128(1997).
DR      EMBL; AF001225; AAB60872.1; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0006955; P:immune response; IEA.
DR      InterPro; IPR000973; CD4_TcAg.
DR      InterPro; IPR007110; IG_I-like.
DR      InterPro; IPR003596; IG_v.
DR      Pfam; PF00047; IG_2.
DR      PRINTS; PR00692; CD4TCANTIGEN.
DR      SMART; SM00406; IG_v.1.
DR      SMART; SM00406; IG_v.1.

```

```

DR      PROSITE; PS50835; IG_LIKE; 1.
FT      NON_TER      1
FT      NON_TER      397
SQ      SEQUENCE      397 AA; 43945 MW; 80C963B92A866CD3 CRC64;

```

## Query Match

```

Best Local Similarity 73.8%; Score 1715; DB 6; Length 397;
Best Local Similarity 89.5%; Pred. No. 1.9e-123;
Matches 332; Conservative 15; Mismatches 24; Indels 0; Gaps 0;

```

```

QY      28 VILGKGGTVELTCTASQKKSIOFWKNSNQIKLNGSGFLTKGPKSLNPAASRLSIW 87
      |||
      1 VILGKGGTVELTCTASQNTTQFWKNSNQIKLNGSGFLTKGSKLRIDSRKSIW 60
DB
QY      88 DQGNPLIITKLIKIDSDTVICEVEDQEEVQLVFGVLTANSDDHLLQGOSLTLLESPP 147
      |||
      61 DQGNPLIITKLIKIDSETTYICEVENKKEBELVFGVLTANSDDHLLQGOSLTLLESPP 120
DB
QY      148 GSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGVTCTVLQNKKEFKIDIVLAFQKA 207
      |||
      121 GSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGVTCTVLQNKKEFKIDIVLAFQKA 180
DB
QY      208 SSIYKKEGEVVEFPFLAFVTEKLTGSGELMWQERASSKSWITFDLKNKEVVKVKT 267
      |||
      181 SSIYKKEGEVVEFPFLAFVTEKLTGSGELMWQERASSKSWITFDLKNKEVVKVKT 240
DB
QY      268 QDPKIQMGKTLPLHITLPOALPOVAGSGNLTALFAKTKLHVEVNLVVMRATQLOKNI 327
      |||
      241 QDPKIQMGKTLPLHITLPOALPOVAGSGNLTALFAKTKLHVEVNLVVMRATQLOKNI 300
DB
QY      328 CEVWPTSPKMLSLIKENKAKVSKREKPVVNLNPEAGMOCLLSDSGQVLLSNIKVL 387
      |||
      301 CEVWPTSPKMLSLIKENKAKVSKREKPVVNLNPEAGMOCLLSDSGQVLLSNIKVL 360
DB
QY      368 PTWSTPVHPRA 398
      |||
      361 PTWSTPVHPMA 371
DB

```

## RESULT 2

```

009260      PRELIMINARY;      PRT;      397 AA.
AC      009260;
DT      01-JUL-1997 (TrEMBLrel. 04, Created)
DT      01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DE      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
GN      CD4 (Fragment).
OS      Cercopithecus sabaenus.
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC      Cercopithecinae; Cercopithecus.
OX      NCB1_TaxID=60711;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=98017879; PubMed=9379478;
RA      Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Machiot C.,
RA      Corbet S., Barre-Sinoussi F., Allan J.S.;
RA      "Relation between phylogeny of African green monkey CD4 genes and
RA      their respective simian immunodeficiency virus genes.";
RL      J. Med. Primatol. 26:120-128(1997).
DR      EMBL; AF001224; AAB60871.1; -.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0016020; C:membrane; IEA.
DR      GO; GO:0006955; P:immune response; IEA.
DR      InterPro; IPR000973; CD4_TcAg.
DR      InterPro; IPR007110; IG_I-like.
DR      InterPro; IPR003596; IG_v.
DR      Pfam; PF00047; IG_2.
DR      PRINTS; PR00692; CD4TCANTIGEN.
DR      SMART; SM00406; IG_v.1.
DR      PROSITE; PS50835; IG_LIKE; 1.
FT      NON_TER      1
FT      NON_TER      397

```

```

SQ SEQUENCE 397 AA; 43882 MW; 478BB277E92EE89 CRC64;
Query Match 73.7%; Score 1713; DB 6; Length 397;
Best Local Similarity 80.2%; Pred. No. 2,7e-123;
Matches 341; Conservative 17; Mismatches 37; Indels 30; Gaps 2;

QY 28 VLGGKGDVTELCTASOKKSIQFHWKNSNQIKILGQGSFLTKGPKSLNDRADSRSLW 87
DB 1 VLGGKGDVTELCTASOKKTTTQPHWKNNSQIKILGQGSFLTKGSKLDRIDSRKSLW 60

QY 88 DQGNFPLIIKNLKIEDSDYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTTLTLESP 147
DB 61 DQCFSMIINKLKIEDSEYICEVENKEEVELLVFGLTANSDTHLLQGQSLTTLTLESP 120

QY 148 GSSPSVQCSPPKRNIOGKRTLSVSOLELQDSGTWCTVYQONKVEFKIDIVVLAFOKA 207
DB 121 GSSPSVQCSPPKRNIOGKRTLSVPOLEQDSGTWCTVYQONKVEFKIDIVVLAFOKA 180

QY 208 SSIVYKKEGQVEFSFPLAFVTEKLTGSGELMWQAEASSSKSMITFDLKNKEVSVKRYT 267
DB 181 SSIVYKKEGQVEFSFPLAFVTEKLTGSGELMWQAEASSSKSMITFDLKNKEVSVKRYT 240

QY 268 QDEKLGWKKLPLHLTPQALPOYAGSGNLTALAEKTKLHQEVNLVVMRATQOENLT 327
DB 241 QDEKLGWKKLPLHLTPQALPOYAGSGNLTALAEKTKLHQEVNLVVMRATQOENLT 300

QY 328 CEVWGPTSPKLMSTLKENKAQVSKREKPVWVNLNPAAGMOCCLSDSGVLLSENIKYL 387
DB 301 CEVWGPTSPKLMSTLKENKAATVSKQAKAVWVNLNPAAGMOCCLSDSGVLLSENIKYL 360

QY 388 PTWSTVHPRASALPAPPTGSALPDPTASALPDPPASALPALAVISFLGLGCV-A 446
DB 361 PTWPTVPQ-----MALIVGVAAGLLFTGLGIFL 391

QY 447 CVLAR 451
DB 392 CVRCR 396

RESULT 3
Q95NE9 PRELIMINARY; PRT; 397 AA.
ID 095NE9
AC 095NE9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CD4 (Fragment).
OS Cercopithecus pygerythrus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=60710;
RN [1]
RP SEQUENCE FROM N.A.
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL: AF001227; AAB60874.1; -.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0006955; P:immune response; IEA.
DR InterPro: IPR000973; CD4_TcAg.
DR InterPro: IPR007110; IG_Like.
DR Pfam: PF00047; IG_2.
DR PRINTS: PR00692; CD4TCANTIGEN.
DR SMART: SM00406; IG_1.
DR PROSITE: PSS0835; IG_Like; 1.
FT NON_TER 1 1
FT 397 397
```

```

SQ SEQUENCE 397 AA; 43946 MW; 21C3E3082ABFC0 CRC64;
Query Match 73.6%; Score 1711; DB 6; Length 397;
Best Local Similarity 80.2%; Pred. No. 3.8e-123;
Matches 341; Conservative 16; Mismatches 38; Indels 30; Gaps 2;

QY 28 VLGGKGDVTELCTASOKKSIQFHWKNSNQIKILGQGSFLTKGPKSLNDRADSRSLW 87
DB 1 VLGGKGDVTELCTASOKKTTTQPHWKNNSQIKILGQGSFLTKGSKLDRIDSRKSLW 60

QY 88 DQGNFPLIIKNLKIEDSDYICEVEDQKEEVQLLVFGLTANSDTHLLQGQSLTTLTLESP 147
DB 61 DQCFSMIINKLKIEDSEYICEVENKEEVELLVFGLTANSDTHLLQGQSLTTLTLESP 120

QY 148 GSSPSVQCSPPKRNIOGKRTLSVSOLELQDSGTWCTVYQONKVEFKIDIVVLAFOKA 207
DB 121 GSSPSVQCSPPKRNIOGKRTLSVPOLEQDSGTWCTVYQONKVEFKIDIVVLAFOKA 180

QY 208 SSIVYKKEGQVEFSFPLAFVTEKLTGSGELMWQAEASSSKSMITFDLKNKEVSVKRYT 267
DB 181 SSIVYKKEGQVEFSFPLAFVTEKLTGSGELMWQAEASSSKSMITFDLKNKEVSVKRYT 240

QY 268 QDEKLGWKKLPLHLTPQALPOYAGSGNLTALAEKTKLHQEVNLVVMRATQOENLT 327
DB 241 QDEKLGWKKLPLHLTPQALPOYAGSGNLTALAEKTKLHQEVNLVVMRATQOENLT 300

QY 328 CEVWGPTSPKLMSTLKENKAQVSKREKPVWVNLNPAAGMOCCLSDSGVLLSENIKYL 387
DB 301 CEVWGPTSPKLMSTLKENKAATVSKQAKAVWVNLNPAAGMOCCLSDSGVLLSENIKYL 360

QY 388 PTWSTVHPRASALPAPPTGSALPDPTASALPDPPASALPALAVISFLGLGCV-A 446
DB 361 PTWPTVPQ-----MALIVGVAAGLLFTGLGIFL 391

QY 447 CVLAR 451
DB 392 CVRCR 396

RESULT 4
O09259 PRELIMINARY; PRT; 397 AA.
ID O09259
AC O09259;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CD4 (Fragment).
OS Cercopithecus sabaues.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=60711;
RN [1]
RP SEQUENCE FROM N.A.
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL: AF001223; AAB60870.1; -.
DR HSP: P01730; IWIQ.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0006955; P:immune response; IEA.
DR InterPro: IPR000973; CD4_TcAg.
DR InterPro: IPR007110; IG_Like.
DR Pfam: PF00047; IG_2.
DR PRINTS: PR00692; CD4TCANTIGEN.
DR SMART: SM00406; IG_1.
DR PROSITE: PSS0835; IG_Like; 1.
FT NON_TER 1 1
```

FT	NON TERM	397	397	
SEQUENCE	397 AA;	43881 MW;	7C39AD0F8506C81 CRC64;	
Query Match		73.5%;	Score 1708;	DB 6; Length 397;
Best Local Similarity		80.0%;	Pred. No. 6,5e-123;	
Matches	340; Conservative	18;	Mismatches 37;	Indels 30; Gaps 2
QY	28	VVLGKKGTVELTCTASQKSKIQFHWKNSNQIKLGNQGSFLTGPKSKLNDRAISRSLM	87	
DB	1	VVLGKKGTVELTCTASQKSKITTTTQTFHWKNSNQIKLGNQGSFLTGPKSKLNDRAISRSLM	60	
QY	88	DQGNPLLIKNLKIEDSPYICEVEDQKEVQLTVFGLTANGDTHLQGSITLTLESPP	147	
DB	61	DQGCSSMIILKNLKIEDSETYICEVENKEEVELVFGLTANSDDHLQGSITLTLESPP	120	
QY	148	GSSPEVQCRSPRGKNIQGGKTLVSQLEQDSGTWCTVYLQNKVKEKIDIVLAFQKA	207	
DB	121	GSSPEVQCRSPRGKNIQGGKTLVSQLEQDSGTWCTVYLQNKVKEKIDIVLAFQKA	180	
QY	208	SSIYVKKGEQVEFSFPPLAFVYEKLTGSGELMWQERASSSKSWITFDLKNREVSRYVT	267	
DB	181	SSTVYKKGEQVEFSFPPLAFVYEKLTGSGELMWQERASSSKSWITFDLKNREVSRYVT	240	
QY	268	QDPKIQMKCKPLPHITLTPALPOYAGSGNLTALAEKTKLHGVNLVYMRATQOXNT	327	
DB	241	QDPKIQMKCKPLPHITLTPALPOYAGSGNLTALAEKTKLHGVNLVYMRATQOXNT	300	
QY	328	CEWNGPTSPKMLSLKLENKEAKVSKREKPVWVWLNPEAGMOCILSDSGVLLSNIKYL	387	
DB	301	CEWNGPTSPKMLSLKLENKATVSKQAKAVVWLNPEAGMOCILSDSGVLLSNIKYL	360	
QY	388	PTWSTPVHPRASALPAPPTGSLPDPQTASALPDPPASALPALAVISFLGLGIGV-A	446	
DB	361	PTWSTPVHPRASALPAPPTGSLPDPQTASALPDPPASALPALAVISFLGLGIGV-A	391	
QY	447	CYLAR 451		
DB	392	CVRCR 396		
RESULT 5				
009263				
AC	009263	PRELIMINARY;	PRT;	397 AA.
DT	01-JUL-1997	(TREMBLrel. 04, Created)		
DT	01-JUL-1997	(TREMBLrel. 04, Last sequence update)		
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)		
DE	CD4 (Fragment).			
GN	CD4.			
OS	Cercopithecus tantalus.			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;			
CC	Cercopithecinae; Cercopithecus.			
CK	NCBI_Taxid=60712;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98017679; PubMed=9379478;			
RA	Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Machiot C.,			
RA	Corbet S., Barre-Sinoussi F., Allan J.S.;			
RT	"Relation between phylogeny of African green monkey CD4 genes and			
RT	their respective simlan immunodeficiency virus genes."			
RL	J. Med. Primatol. 26:120-128 (1997).			
DR	EMBL; AF001222; AAB60869.1; -.			
DR	HSSP; P01730; 1WIO.			
DR	GO; GO:0016020; Cmembrane; IEA.			
DR	GO; GO:0006955; P:immune response; IEA.			
DR	InterPro; IPR000973; CD4_TCSG.			
DR	InterPro; IPR007110; IG_1like.			
DR	InterPro; IPR003596; IG_V.			
DR	Pfam; PF00047; ig_2			
DR	PRINTS; PR00692; CD4TCANTIGEN.			
DR	SMART; SM00406; IGV_1.			
DR	PROSITE; PS50835; IG_LIKE; 1.			

FT	NON TER	1	1
FT	SEQUENCE	397 AA; 43994 MW; A3CD031535A51524 CRC64;	
SO	SEQUENCE	397 AA; 43994 MW; A3CD031535A51524 CRC64;	
	Query Match	73.4%; Score 1706; DB 6; Length 397;	
	Best Local Similarity	88.9%; Pred. No. 9,3e-123;	
	Matches 330; Conservative 15; Mismatches 26; Indels 0; Gaps 0;		
OY		28 VLICKGGDTVELTCTASQKKSIOFHWNQNSQIKILGNQGSFLLTKGSPSKLNDRADSRSLW 87	
DB		1 VLAKKGGDTVELTCTASQNTTTFHWNQNSQIKILGNQGSFLLTKGSSKLRDRIDSRSLW 60	
OY		88 DQGNFPLILIKNLKLEDDDTYICEDDQKEFVQLLVFGLTANSDPHLLQGSLLTLTLSEPP 147	
DB		61 DQGFSEMIKNLKLIEDDETYICEVENKKEVEYELLVFGLTANSDPHLLQGSLLTLTLSEPP 120	
OY		148 GSSPSVOCRSPRGNNIOGGKTLASOLELDQSGTWCTVLONOKKVEFKIDIVLAFOKA 207	
DB		121 GSSPSVCRSPRGNNIOGGKTLASPOLERPDQSGTWCTVSDQNTVEFKIDIVLAFOKA 180	
OY		208 SSIYVYKKEGEQVEFSPLAFTVEKLITSGSELWQAERASSKSMITFDLKNKEVSVKRYT 267	
DB		181 SSTYVYKKEGEQVEFSPLAFTVEKLITSGSELWQAERASSKSMITFDLKNKEVSVKRYT 240	
OY		268 QDPKLQNGKPLPLHLITPOLPQVAGSGNLTALFAETGKGLHGVNLVYMPATQLOKULT 327	
DB		241 QDPKLQNGKPLPLHLITPOLPQVAGSGNLTALFAETGKGLHGVNLVYMPATQFOEHLT 300	
OY		328 CEVWGPSPKMLSLKLENKAKYSSKKEKPYVWVNLPEAGNMQCLLSDSGVYLLSNIKYL 387	
DB		301 CEVWGPSPKMLSLKLENKAKYSSKQAKAVWVNLPEAGNMQCLLSDSGVYLLSNIKYL 360	
OY		388 PTWSTPYHPRA 398	
DB		361 PTWPTPVQPM 371	
RESULT 6			
009262			
ID	009262	PRELIMINARY;	PRT; 397 AA.
AC	009262;		
DT	01-JUL-1997	(TREMBLrel. 04, Created)	
DT	01-JUL-1997	(TREMBLrel. 04, Last sequence update)	
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)	
DE	CD4 (Fragment).		
GN	CD4.		
OS	Cercopithecus tantalus.		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;		
OC	Cercopithecinae; Cercopithecus.		
OX	NCBI_TaxID=60712;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=98017879; PubMed=9379478;		
RA	Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,		
RA	Corbet S., Baire-Shinubei F., Allan J.S.;		
RT	"Relation between phylogeny of African green monkey CD4 genes and		
RT	their respective simian immunodeficiency virus genes.";		
RL	J. Med. Primatol. 26:120-128(1997).		
DR	EMBL; AF001221; AB60868.1; -.		
DR	HSSP; P01730; 1WQ.		
DR	GO; GO:0016020; C:membrane; IEA.		
DR	GO; GO:0006955; P:immune response; IEA.		
DR	InterPro; IPR000973; CD4_TGAG.		
DR	InterPro; IPR007110; IG_V.		
DR	InterPro; IPR003596; IG_V.		
DR	Pfam; PF00047; IG_2.		
DR	PRINTS; PRO0692; CD4TCANTIGEN.		
DR	SMART; SMO0406; IGV. 1.		
DR	PROSITE; PS50835; IG_LIKE. 1.		
FT	NON_TER	1	1
FT	NON_TER	397	397
SO	SEQUENCE	397 AA; 43954 MW; CE7PF5D82335B0D CRC64;	

Query Match	73.1%	Score 1699	DB 6	Length 397
Beet Local Similarity	88.9%	Pred. No. 3,2e-122		
Matches 330	Conservative 15	Mismatches 26	Indels 0	Gaps 0
QY	28	VLAGKGGTVLTCASOKKSIQFMKNSNOIKILNGSFFLTQPSKILNDRADRSRLM	87	
DB	1	VLAGKGGTVLTCASQNTTTQFMKNSNOIKILNGSFLTKSSKLRIRISRSKLW	60	
QY	88	DQGNFPLIIKNIKIEDSDTYICEVDOKEBQVLVFGLTANSDPHLLGOSITLTLESPP	147	
DB	61	DQGCMSMIKNIKIDSEYIICEVKNKKEVELLVFGLTANSDPHLLGOSITLTLESPP	120	
QY	148	GSSPEVOCRSPRKNIQGGKTLVSQLELOBSGTCTVYNQKKVEPKIIVLAAPQA	207	
DB	121	GSSPEVCKRSPRGKNIQGGRTLSVQLEKRODSGTCTCTVSQDNVEXKIDIVLAAPQA	180	
QY	208	SSIIVKKEGEOVESFPPLAFIVEKLTGSGELMWQERASSSKSWITFDPLKNKEVSKRYT	267	
DB	181	SSIIVKKEGEOVEFSPPLAFITLEKLTGSGELMWQERASSSKSWITFDPLKNKEVSKQYT	240	
QY	268	ODPKLQWKKPLPHHTLPQALPOVYGSNLTALTEPAKTKGKHQENLVAMRATQOKYLT	327	
DB	241	QDPFLQWKKPLPHHTLPQALPOVYGSNLTALTEPAKTKGKHQENLVAMRATQOQENLYT	300	
QY	338	CEVMGPTSPKMLSLIKLENKEAKSKREKPVVNLNPEAGMOCULSDSGOVLLESNIIKYL	387	
DB	301	CEVMGPTSPKMLSLIKLENKAATVATSKQAKAVVNLNPEAGMOCULSDSGOVLLESNIIKYL	360	
QY	388	PTWSTPVHPRA 398		
DB	361	PTWPTPVQPMNA 371		

```

RESULT 7
ID 002805 PRELIMINARY; PRT: 397 AA.
AC 002805; 077593;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE (Fragment).
GN CD4.
OS Cercopithecus aethiops (Green monkey) (Givet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopithecinae; Cercopithecus.
NCBI_TaxID=9534;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Peripheral blood;
RC MEDLINE=98017879; Pubmed=9379478;
RX Fomgaard A., Mueller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128 (1997).
[2]
RN SEQUENCE OF 80-165 FROM N.A.
RP MEDLINE=98320644; Pubmed=9656488;
RA Harris E.E., Diocetti T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabays (Primates: Papionini)";
RL Mol. Biol. Evol. 15:892-900 (1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH p56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTRAINS
CC ONE V-LIKE AND 3 C2-LIKE DOMAINS.
DR EMBL; AF001226; AAB60873.1; -.

```

DR	EMBL: AF057380; AAC25124.1; -.	
DR	HSSP: P01730; IMQ.	
DR	GO: GO:0016021; C: integral to membrane; IEA.	
DR	GO: GO:0006955; P: immune response; IEA.	
DR	InterPro: IPR000973; CD4_TcRg.	
DR	InterPro: IPR007110; Ig-like.	
DR	InterPro: IPR003596; Ig_v.	
DR	Pfam: PF00047; Ig_2.	
DR	PRINTS: PR00692; CD4TCANTIGEN.	
DR	SMART; SM00406; IGV; 1.	
KW	Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell; Lipoprotein; Palmitate; Repeat.	
KW	NON_TER	1
FT	DOMAIN	<1 370
FT	TRANSMEM	371 391
FT	DOMAIN	392 >397
FT	DOMAIN	<1 98
FT	DOMAIN	99 176
FT	DOMAIN	177 290
FT	DOMAIN	291 347
FT	CARBOHYD	15 15
FT	CARBOHYD	30 30
FT	CARBOHYD	269 269
FT	CARBOHYD	298 298
FT	DISULFID	14 82
FT	DISULFID	128 157
FT	DISULFID	301 343
FT	LIPID	392 392
FT	LIPID	395 395
FT	NON_TER	397 397
SEQ	SEQUENCE	397 AA; 43980 MW; F74CA2E322B196155 CRC64;

Query Match	73.1%	Score 1699;	DB 6;	Length 397;
Best Local Similarity	79.8%;	Pred. No. 3,28-122;		
Matches	339;	Conservative	16;	Mismatches 40; Indels 30; Gaps 2;
QY	28	VILAKGQDYVELTCTASQKKSIOFHWNKSNQIKLGNQGSFLTKGSPSKLNDRADRSRLW	87	
Db	1	VLLAKKGQDYVELTCTNASQKTTTQPHMNKSNQTIKLGKQGSFLTKGSSSKLDRIDSRKSLW	60	
QY	88	DQGFPLILINKLKTEDSDTYICEVEDQKEVQULVPELTANSPTHLLQGSSTLTLESPP	147	
Db	61	DQGFPSMILINKLKTEDSETYICEVENKKEVELLVPFLTANSPTHLLQGSSTLTLESPP	120	
QY	148	GSSSPVOCRSPRGKNIQGGKTLISVQLELQDSGTWTCTVLQNKKEVFKLIDIVLAFQKA	207	
Db	121	GSSSPVACRSPRGKNIQGGKTLISVQLELQDSGTWTCTVLQNKKEVFKLIDIVLAFQKA	180	
QY	208	SSIIVYKKEGEQVEFSPPLAFTVEKLTGSGELMWQAEARASSKSWITFDLKNKEVSVKVT	267	
Db	181	SSIIVYKKEGEQVEFSPPLAFTVEKLTGSGELMWQAEARASSKSWITFDLKNKEVSVKVT	240	
QY	268	QDPKLQNGKKLPLHLTPQALPOYAGSNTLTALBATGKLHQEVLVYVRAIQLOQKNT	327	
Db	241	QDPKLQNGKKLPLHLTPQALPOYAGSNTLTALBATGKLHQEVLVYVRAIQLOQKNT	300	
QY	328	CEVWGPSPKLMSTLKENKEAKVSKSEKVVYVNLNPAAGMOCILSDSGOVLLESNTKVL	387	
Db	301	CEVWGPSPKLMSTLKENKEAKVSKSEKVVYVNLNPAAGMOCILSDSGOVLLESNTKVL	360	
QY	388	PTWSTPVHPRAASALPAPEPTGSALPDQOTASALPDPPAASALPALAAVISFLIGLIGLV-A	446	
Db	361	PTWSTPVHPRAASALPAPEPTGSALPDQOTASALPDPPAASALPALAAVISFLIGLIGLV-A	391	
QY	447	CVLAR 451		
Db	392	CVRCR 396		

RESULT 8

Q8H2T8      PREDIMINARY;      PRT;      457 AA.

AC    Q8H2T8;      457 AA.

DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Lymphocyte antigen CD4.  
 OS Callicthrix jacchus (Common marmoset).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callitrich.  
 NCBI\_TaxID=9483;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22174698; PubMed=12186836;  
 RA Labonte J.A., Babcock G.J., Patel T., Sodroski J.;  
 RT "Blockade of HIV-1 Infection of New World Monkey Cells Occurs  
 Primarily at the Stage of Virus Entry."  
 RL J. Exp. Med. 196:431-445(2002).  
 DR EMBL; AF452616; AAN14532.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0006955; P:immune response; IEA.  
 DR InterPro; IPR000973; CD4\_TcAg.  
 DR InterPro; IPR003599; IG.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; IG\_2.  
 DR PRINTS; PR00692; CD4TCANTIGEN.  
 DR SMART; SM00409; IG; 3.  
 DR SMART; SM00406; IGv; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 SQ SEQUENCE 457 AA; 50878 MW; 718CDB78D97F59B CRC64;

Query Match 69.5%; Score 1615.5; DB 6; Length 457;  
 Best Local Similarity 70.7%; Pred. No. 1e-115;  
 Matches 324; Conservative 37; Mismatches 58; Indels 39; Gaps 4;

QY 1 MNRGVPFRHLVLTQALLPAATQGNKVVYLGKKGDVETLTCTASOKKSIQFHMKNNOIK 60  
 DB 1 MNGGIPFRHLVLTQALLPAATQGNKVVYLGKKGEVETLTCTASOKKSIQFHMKNNOIK 60  
 QY 61 ILGNOSFLTQPSKLNDRADSRSLMDQGNPPLIKNLKIEDSDTYICEVEDOKEEVL 120  
 DB 61 ILGIGSFYTKGQSKLANRIDSQSWDRSPFLIRNVQVESEYICEVESKKEEVL 120  
 QY 121 LVFGLTANSDTHLQGOSSLTLESPPGSSPVQCSPPGKNIQGGKTLSVSOLELDSDG 180  
 DB 121 QVFGTLVNPDTHLQGOSSLTLESPPGSSPVQCSPPGKNIQGGKTLSVSOLELDSDG 180  
 QY 121 QVFGTLVNPDTHLQGOSSLTLESPPGSSPVQCSPPGKNIQGGKTLSVSOLELDSDG 180  
 DB 121 QVFGTLVNPDTHLQGOSSLTLESPPGSSPVQCSPPGKNIQGGKTLSVSOLELDSDG 180  
 QY 181 TWCTVLTQNKVEPKIDIVLAFQKASSIVYKKEGEVETLTCTASOKKSIQFHMKNNOIK 240  
 DB 181 TWCTVLTQNKVEPKIDIVLAFQKASSIVYKKEGEVETLTCTASOKKSIQFHMKNNOIK 240  
 QY 181 TWCTVLTQNKVEPKIDIVLAFQKASSIVYKKEGEVETLTCTASOKKSIQFHMKNNOIK 240  
 DB 181 TWCTVLTQNKVEPKIDIVLAFQKASSIVYKKEGEVETLTCTASOKKSIQFHMKNNOIK 240  
 QY 241 QAERASSSKSWITFDLKNKEVSVKRVYDQPKLQMGKKLPLHLTLPOALPOYAGSGNLTIA 300  
 DB 240 QAERASSSKSWITFDLKNKEVSVKRVYDQPKLQMGKKLPLHLTLPOALPOYAGSGNLTIA 300  
 QY 301 LEAKTGKLEHVEVNLVVMRATQLOKNTLCEVWGPTSPKMLSLKENKEAVSKREKPVVW 360  
 DB 300 LKKTGTGLHVEVNLVVMRATQLOKNTLCEVWGPTSPKMLSLKENKEAVSKREKPVVW 360  
 QY 361 LNPBAGMOCCLSDSGOVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420  
 DB 360 LNPBAGMOCCLSDSGOVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420  
 QY 421 DPPAASALPALAVISFLGL-----GLGV-ACVLAART 453  
 DB 395 -----PVALIVLGVAAGLAVFTGLGIFLCVRCRR 424

RESULT 9  
 Q8H2T7 PRELIMINARY; PRT; 457 AA.  
 AC Q8H2T7;  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE Lymphocyte antigen CD4.  
 OS Saimiri sciureus (Common squirrel monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae; Saimiri.  
 NCBI\_TaxID=9521;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22174698; PubMed=12186836;  
 RA Labonte J.A., Babcock G.J., Patel T., Sodroski J.;  
 RT "Blockade of HIV-1 Infection of New World Monkey Cells Occurs  
 Primarily at the Stage of Virus Entry."  
 RL J. Exp. Med. 196:431-445(2002).  
 DR EMBL; AF452617; AAN14533.1; -.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0006955; P:immune response; IEA.  
 DR InterPro; IPR000973; CD4\_TcAg.  
 DR InterPro; IPR003599; IG.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003596; IG\_v.  
 DR Pfam; PF00047; IG\_2.  
 DR PRINTS; PR00692; CD4TCANTIGEN.  
 DR SMART; SM00409; IG; 3.  
 DR SMART; SM00406; IGv; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 1.  
 SQ SEQUENCE 457 AA; 50899 MW; B164DA89E70C575A CRC64;

Query Match 68.2%; Score 1585.5; DB 6; Length 457;  
 Best Local Similarity 70.1%; Pred. No. 2e-113;  
 Matches 321; Conservative 34; Mismatches 64; Indels 39; Gaps 4;

QY 1 MNRGVPFRHLVLTQALLPAATQGNKVVYLGKKGDVETLTCTASOKKSIQFHMKNNOIK 60  
 DB 1 MNGGIPFRHLVLTQALLPAATQGNKVVYLGKKGEVETLTCTASOKKSIQFHMKNNOIK 60  
 QY 61 ILGNOSFLTQPSKLNDRADSRSLMDQGNPPLIKNLKIEDSDTYICEVEDOKEEVL 120  
 DB 61 ILGIGSFYTKGQSKLANRIDSQSWDRSPFLIRNVQVESEYICEVESKKEEVL 120  
 QY 121 LVFGLTANSDTHLQGOSSLTLESPPGSSPVQCSPPGKNIQGGKTLSVSOLELDSDG 180  
 DB 121 QVFGTLVNPDTHLQGOSSLTLESPPGSSPVQCSPPGKNIQGGKTLSVSOLELDSDG 180  
 QY 121 QVFGTLVNPDTHLQGOSSLTLESPPGSSPVQCSPPGKNIQGGKTLSVSOLELDSDG 180  
 DB 121 QVFGTLVNPDTHLQGOSSLTLESPPGSSPVQCSPPGKNIQGGKTLSVSOLELDSDG 180  
 QY 181 TWCTVLTQNKVEPKIDIVLAFQKASSIVYKKEGEVETLTCTASOKKSIQFHMKNNOIK 240  
 DB 181 TWCTVLTQNKVEPKIDIVLAFQKASSIVYKKEGEVETLTCTASOKKSIQFHMKNNOIK 240  
 QY 181 TWCTVLTQNKVEPKIDIVLAFQKASSIVYKKEGEVETLTCTASOKKSIQFHMKNNOIK 240  
 DB 181 TWCTVLTQNKVEPKIDIVLAFQKASSIVYKKEGEVETLTCTASOKKSIQFHMKNNOIK 240  
 QY 241 QAERASSSKSWITFDLKNKEVSVKRVYDQPKLQMGKKLPLHLTLPOALPOYAGSGNLTIA 300  
 DB 240 QAERASSSKSWITFDLKNKEVSVKRVYDQPKLQMGKKLPLHLTLPOALPOYAGSGNLTIA 300  
 QY 301 LEAKTGKLEHVEVNLVVMRATQLOKNTLCEVWGPTSPKMLSLKENKEAVSKREKPVVW 360  
 DB 300 LKKTGTGLHVEVNLVVMRATQLOKNTLCEVWGPTSPKMLSLKENKEAVSKREKPVVW 360  
 QY 361 LNPBAGMOCCLSDSGOVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420  
 DB 360 LNPBAGMOCCLSDSGOVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420  
 QY 421 DPPAASALPALAVISFLGL-----GLGV-ACVLAART 453  
 DB 395 -----PVALIVLGVAAGLAVFTGLGIFLCVRCRR 424

RESULT 10  
 Q9XS78 PRELIMINARY; PRT; 455 AA.  
 AC Q9XS78;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE T-cell surface glycoprotein CD4.  
 OS Delphinapterus leucas (beluga whale).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti;
OC Monodontidae; Delphinapterus.
OX NCBI_TaxId=9749;
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Thymus;
RX MEDLINE=99216435; PubMed=10199913;
RA Romano T.A., Ridgway S.H., Felten D.L., Quaranta V.,
RT "Molecular cloning and characterization of CD4 in an aquatic mammal,
RL the white whale Delphinapterus leucas."
DR EMBL; AF071799; AAD23738.1; -.
DR HSSP; P01730; 1WIO.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TcRg.
DR InterPro; IPR007110; Ig_I-like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_3.
DR PRINTS; PR00692; CD4TCRNTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 2.
SQ SEQUENCE 455 AA; 50499 MW; AA532FD411AASD1 CRC64;

Query Match 51.1%; Score 1189; DB 6; Length 455;
Best Local Similarity 56.4%; Pred. No. 7e-83;
Matches 257; Conservative 48; Mismatches 105; Indels 46; Gaps 6;

QY 1 MNRGVPFRHLVLVQLALPAAATQGNKRVLGKKGDTVELTCTASQKKSIQFHWKNSNOIK 60
DB 1 MDPRTSLRHLFLVLQVLMPLAGTGKRVLGKAGELALPCQASQKKYMTFTWRLSSQV 60
QY 61 ILGNQGFLLTKGPKLNDRADRSRLMDQGNFPIIKNLKIEDSDTYICEVEDQKEEV 120
DB 61 ILGNHGFYFMHKGASNLHSRVESKINLMDQSFPLVIDLEVPDGTIYICEVEDKKEVEL 120
QY 121 LVFGLTANSPTHLLQGGSLTLTLESPGSSPSVQCRSPRGNIQGGKTLVSQLELDQSG 180
DB 121 QVRLTLNASSDTRLNGSLTLTLESPGSSNPSPVQMKPRGNKRNKAESLSPQVGLDQSG 180
QY 181 TWTCTVLQONKKEVFKIDIVLAFQKASIVYKKEGEQVEFSPFLAFTEKLTGSGEL-W 239
DB 181 TWTCTVQAQQTIVFNHGLIIVLAFQEVSTVYAKEGQNMPSFLVTGDBNL--SGSLSW 238
QY 240 WQERASSSSKSWITFDLKNKEVSVKRVTDQPKLQMKKLPULHTLPALQYVAGSGNLT 239
DB 240 WQERASSSSKSWITFDLKNKEVSVKRVTDQPKLQMKKLPULHTLPALQYVAGSGNLT 239
QY 239 LQAKGSSPSWSWITFDLKNKEVSVKRVTDQPKLQMKKLPULHTLPALQYVAGSGNLT 238
DB 300 ALKAKTGKLEHGVNLYVMRATQLOKNTCEVWGPTSPKLMLSLKEKKAQVSRKRPVW 359
DB 299 NL--TKGKLYQEVNLYVMRATQLOKNTCEVWGPTSPKLMLSLKEKKAQVSRKRPVW 356
QY 360 VLNPEAGMOCCLSDSGQVLLLESNIKVLPTWSTPVRHRSALPAPPTGSALPPQQTASAL 419
DB 357 VLGEAGMOCCLSDSGQVLLLESNIKVLPTWSTPVRHRSALPAPPTGSALPPQQTASAL 419
QY 420 PDPPASALPAALAVT-----SFLIGLGLVACVLA 450
DB 386 --PVLAAHWPKLLAVLVGIGTISLLLAGF---CIFSS 416

RESULT 11
ID P79355 PRELIMINARY; PRT; 474 AA.
AC P79355;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CD4 antigen precursor.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxId=9685;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RA Miyazawa T.,
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92165316; PubMed=1537604;
RA Noriine J., Miyazawa T., Kawaguchi Y., Tohya Y., Kai C., Mikami T.;
RT "A cDNA encoding feline CD4 has a unique repeat sequence downstream of
RL the V-like region."
DR EMBL; AB000483; BA19124.1; -.
DR HSSP; P01730; 1WIO.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TcRg.
DR InterPro; IPR007110; Ig_I-like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_3.
DR PRINTS; PR00692; CD4TCRNTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 2.
KM SIGNAL.
FT CHAIN 1 26 POTENTIAL.
SQ SEQUENCE 474 AA; 52243 MW; D946DDABEAD00EC CRC64;

Query Match 46.8%; Score 1087; DB 6; Length 474;
Best Local Similarity 57.6%; Pred. No. 4.2e-75;
Matches 239; Conservative 49; Mismatches 99; Indels 28; Gaps 6;

QY 1 MNRGVPFRHLVLVQLALPAAATQGNKRVLGKKGDTVELTCTASQKKSIQFHWKNSNOI 59
DB 1 MNGGAVFRLHLVLQVLMPLAAGTGAELPCQASQKKYMTFTWRLSSQV 60
QY 60 KILGNQGFLLTKGPKLNDRADRSRLMDQGNFPIIKNLKIEDSDTYICEVEDQKEEV 118
DB 61 KILSGHSLSLCTLGSSKLTFRPSKILMDQGSFPLVILSLQVADSGITCEVENKREV 120
QY 119 QLLVFGILTANSD-----THLLQGGSLTLTLESPGSSPSVQCRSPRGK 161
DB 121 ELVFGILTAKVDPGSGSGSSSTSTSTSYLLQGGSLTLTLESPASSNPSPVQMKPRGNK 180
QY 162 NIOGKTLVSQLELDQSGTWTCTVLQONKKEVFKIDIVLAFQKASIVYKKEGEQVEF 221
DB 181 SKSGVHSLISQLELDQSGTCTVTSQSKTLVFNNILVLAARKVSNVYAKEGQVEF 240
QY 222 SPFLAFTEKLTGSGELMNOAERASSKSWITFDLKNKEVSVKRVTDQPKLQMKKLPULH 281
DB 241 SPFLANFEDENLKN--LRNKAEGAPSSLMISTTLNKKQLSKVEVDPYSKLQMDSLPLR 298
QY 282 LTLPOALPOVAGSGNLTALAEKTKLHGVNLYVMRATQLOKNTCEVWGPTSPKLMLS 341
DB 299 FTLPVNLISRYAGSGNLTLLVID--KGLOQEVKLVVRATQLOKNTCEVWGPTSPKLTLS 356
QY 342 LKLEKKAQVSRKRPVWVLNPEAGMOCCLSDSGQVLLLESNIKULP-----TWS 391
DB 357 LKLEKKAQVSRKRPVWVLNPEAGMOCCLSDSGQVLLLESNIKULP-----TWS 391
DB 357 LKLEKKAQVSRKRPVWVLNPEAGMOCCLSDSGQVLLLESNIKULP-----TWS 391

RESULT 12
ID Q61396 PRELIMINARY; PRT; 457 AA.
AC Q61396;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE T-cell differentiation antigen.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88041159; PubMed=2823269;
RA Gorman S.D., Tourvieille B., Barnes J.R.,
RT "Structure of the mouse gene encoding CD4 and an unusual transcript in
   brain."
RL Proc. Natl. Acad. Sci. U.S.A. 84:7644-7648(1987).
DR EMBL: M17080; AAA37402.1; -.
DR EMBL: M17076; AAA37402.1; JOINED.
DR EMBL: M17077; AAA37402.1; JOINED.
DR EMBL: M17078; AAA37402.1; JOINED.
DR EMBL: M17079; AAA37402.1; JOINED.
DR HSSP: P01730; 1WBR.
DR MGD: MGI:88335; Cd4.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0006955; P:immune response; IEA.
DR InterPro: IPR000973; CD4_TCRg.
DR InterPro: IPR007110; Ig_Like.
DR InterPro: IPR003596; Ig_v.
DR Pfam: PF00047; Ig_2.
DR PRINTS: PR00692; CD4TCANTIGEN.
DR SMART: SM00406; IGV; 1.
DR PROSITE: PSS0835; IG_LIKE; 1.
SQ SEQUENCE 457 AA; 51368 MW; 24AB19EDA285B5D0 CRC64;

Query Match 42.4%; Score 985; DB 11; Length 457;
Best Local Similarity 53.6%; Pred. No. 2,7e-67;
Matches 220; Conservative 62; Mismatches 110; Indels 10; Gaps 7;

QY 1 MNRGVPRRH-LILVLOLALLPAATQGNKVVLLGKGGDTVEITCTASQKSIQPHMKNNOI 59
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 MCRATISLRLLLLLOLSQLAVTOEKTIVLKGESAEIPCESSQKLTIVFTWKFSDOR 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 60 KILGNG-SFLTGG--PSKLNDRADSRRLMDQGNPLIIKNIKIDSDPTVCEVDDKE 116
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 KILGQHGKGVLLINGSPSQ-DRPDSKKGAMKSGSPFLINLKLMEDSQTICELNRRE 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 117 EVOLVFGLTANSDTHLLQGSITLTLES-PGSSPSVQCRSPRGKNIQSGKTLVSQLE 175
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 120 EVELMWFKYTFSPGTSLQGSITLTLDNSKXSNPLTECKHKKGKGVSSKYLMSNLR 179
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 176 LDGSGTWTCTVLONQKKEFKIDIVLAFOKASSIYKKEGEQVEFSPLAFTVEKLTGS 235
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 180 VQDSDFMNCVTLLDQKNMFGMTLSVLFQSTAITAYKSEGSAAEFSPINFABE--NGW 237
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 236 GELMWOAERASSKSNITPDLKNKEVSVKRVTDPKLQMGKRLPLHLTLPOALPOYAGSG 295
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 238 GELMWAERKDSFPQPMISFISIKNEKSVQKSTQDLKQLKETPLTLTKIPQVSLQFAGSG 297
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 296 NLTLALEAKTGKLEHENVLVVVRATQLOKNLTCEVWGPTSPKLMLSIKLENKEAKYSKRE 355
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 298 NLTLTLD-KGTLHGEENVLVVVRQAOLNNTLTCEVWGPTSPKRLTLTKQENQEARVSEBQ 355
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 356 KPYVAVLNPENAGMQCLISDGOVLLESNIKVL 387
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 356 KVVQVAVABETGLMQCLISDGDKVMDSRIQVL 387
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 13
O55054 PRELIMINARY; PRT; 433 AA.
ID 055054
AC O55054;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE T4 surface glycoprotein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=88097446; PubMed=3501122;
RA Maddon P.J., Molinaux S.M., Maddon D.E., Zimmerman K.A., Godfrey M.,
RA Alt F.W., Chess L., Axel R.;
RT "Structure and expression of the human and mouse T4 genes."
RL Proc. Natl. Acad. Sci. U.S.A. 84:9155-9159(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA Maddon P.J., Molinaux S.M., Maddon D.E., Zimmerman K.A., Godfrey M.,
RA Alt F.W., Chess L., Axel R.;
RL Submitted (FE8-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF045882; AAC01764.1; -.
DR HSSP: P01730; 1WBR.
DR MGD: MGI:88335; Cd4.
DR GO: GO:0016020; C:membrane; IEA.
DR GO: GO:0006955; P:immune response; IEA.
DR InterPro: IPR000973; CD4_TCRg.
DR InterPro: IPR007110; Ig_Like.
DR InterPro: IPR003596; Ig_v.
DR Pfam: PF00047; Ig_2.
DR PRINTS: PR00692; CD4TCANTIGEN.
DR SMART: SM00406; IGV; 1.
DR PROSITE: PSS0835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 433
SQ SEQUENCE 433 AA; 48590 MW; AB19330750A8499A CRC64;

Query Match 41.1%; Score 954.5; DB 11; Length 433;
Best Local Similarity 54.1%; Pred. No. 5.6e-65;
Matches 199; Conservative 59; Mismatches 101; Indels 9; Gaps 6;

QY 24 QGNKVVLLGKGGDTVEITCTASQKSIQPHMKNNOIKILGNG-SFLTGG--PSKLNDR 80
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 QGKTLVLKGESAEIPCESSQKLTIVFTWKFSDRKILGQHGKGVLLINGSPSQ-DRF 59
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 81 DSRSLMDQGNPLIIKNIKIDSDPTVCEVDDKEVQVLVFGLTANSDTHLLQGSITL 140
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 60 DSKKGAMKSGSPFLINLKLMEDSQTICELNRKEVELMWFKYTFSPGTSLQGSITL 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 141 LTLES-PGSSPSVQCRSPRGKNIQSGKTLVSQLELDGSGTWTCTVLONQKKEFKIDI 199
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 120 LTLDNSKXSNPLTECKHKKGKGVSSKYLMSNLRVQDSDFMNCVTLLDQKNMFGMTL 179
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 200 VVLAFOKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMWOAERASSKSNITDCLK 259
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 180 SVLGFQSTAITAYKSEGSAAEFSPINFABE--NGMGEIMWAERKDSFPQPMISFISIK 237
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 260 EYVSVKRVTDPKLQMGKRLPLHLTLPOALPOYAGSGNLTLALEATGKLEHENVLVVVR 319
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 238 EYVSVKSTQDLKQLKETPLTLTKIPQVSLQFAGSGNLTLLD--KGTLHGEENVLVVVR 295
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 320 TQLOKNLTCEVWGPTSPKLMLSIKLENKEAKYSKREKPYVAVLNPENAGMQCLISDGOVL 379
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 296 AQLNNTLTCEVWGPTSPKRLTLTKQENQEARVSEBQKVVQVAVABETGLMQCLISDGD 355
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 380 LESNIKVL 387
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 356 MDSRIQVL 363
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 14
O77596 PRELIMINARY; PRT; 86 AA.
ID 077596
AC O77596;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
   (Fragment).
OS Mandillius sphinx (Mandril) (Papio sphinx).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Mandrillus.
OX NCBI_TaxID=9561;

```



```
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RL mangabeys (primates: Papionini).";
RM Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL: AF057386; AAC25130.1; -.
DR HSSP: P01730; ICYD.
DR GO: GO:0016021; C:Integral to membrane; IEA.
DR InterPro: IPR007110; Ig-like.
KM Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT DOMAIN 1 1
FT DISULFID 20 >86 IG-LIKE V-TYPE DOMAIN.
FT NON TER 49 78 IG-LIKE C2-TYPE DOMAIN.
SQ SEQUENCE 86 AA; 9406 MW; 2BD97A9EE19582AB CRC64;

Query Match 17.2%; Score 400; DB 6; Length 86;
Best Local Similarity 89.4%; Pred. No. 2.3e-23;
Matches 76; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 107 YICEVEDKKEEVLVFGLTANSPTHTLQGOSITLTLESPPGSSPSVQCRSPRGKNIQG 166
DB 1 YICEVEDKKEEVLVFGLTANSPTHTLQGOSITLTLESPPGSSPSVQCRSPRGKNIQG 60
167 KTLSVQLSDSGTWTCTVQLONOK 191
DB 61 RTLSVPQLERQDSGTWCTVQSODQK 85

RESULT 15
QY 077597 PRELIMINARY; PRT; 86 AA.
AC 077597;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
OS Mandrillus leucophaeus (Drill) (Papio leucophaeus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Mandrillus.
OX NCBI_TaxID=9568;
RN 1
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RL mangabeys (primates: Papionini).";
RM Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL: AF057387; AAC25131.1; -.
DR HSSP: P01730; ICYD.
DR GO: GO:0016021; C:Integral to membrane; IEA.
DR InterPro: IPR007110; Ig-like.
KM Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT DOMAIN 1 1
FT DISULFID 20 >86 IG-LIKE V-TYPE DOMAIN.
FT NON TER 49 78 IG-LIKE C2-TYPE DOMAIN.
SQ SEQUENCE 86 AA; 9417 MW; 371CA39EF58182AB CRC64;

Query Match 17.1%; Score 397; DB 6; Length 86;
Best Local Similarity 89.4%; Pred. No. 3.9e-23;
Matches 76; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
```

```
FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.
FT DISULFID 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT NON TER 49 78 BY SIMILARITY.
SQ SEQUENCE 86 AA; 9406 MW; 2BD97A9EE19582AB CRC64;

Query Match 17.2%; Score 400; DB 6; Length 86;
Best Local Similarity 89.4%; Pred. No. 2.3e-23;
Matches 76; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 107 YICEVEDKKEEVLVFGLTANSPTHTLQGOSITLTLESPPGSSPSVQCRSPRGKNIQG 166
DB 1 YICEVEDKKEEVLVFGLTANSPTHTLQGOSITLTLESPPGSSPSVQCRSPRGKNIQG 60
167 KTLSVQLSDSGTWTCTVQLONOK 191
DB 61 RTLSVPQLERQDSGTWCTVQSODQK 85

RESULT 16
QY 077594 PRELIMINARY; PRT; 86 AA.
AC 077594;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
OS (Fragment).
OC Cercopithecus mitis.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=36225;
RN 1
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RL mangabeys (primates: Papionini).";
RM Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL: AF057381; AAC25125.1; -.
DR HSSP: P01730; ICYD.
DR GO: GO:0016021; C:Integral to membrane; IEA.
DR InterPro: IPR007110; Ig-like.
KM Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT DOMAIN 1 1
FT DISULFID 20 >86 IG-LIKE V-TYPE DOMAIN.
FT NON TER 49 78 IG-LIKE C2-TYPE DOMAIN.
SQ SEQUENCE 86 AA; 9417 MW; 371CA39EF58182AB CRC64;

Query Match 17.1%; Score 397; DB 6; Length 86;
Best Local Similarity 89.4%; Pred. No. 3.9e-23;
Matches 76; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
```

RESULT 17



```

Query Match      16.9%; Score 392; DB 6; Length 86;
Best Local Similarity 87.1%; Pred. No. 9.5e-23;
Matches 74; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 107 YICEVEDQKEEVOLVFGLTANSPDTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGG 166
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      1 YICEVEDKKEVELLVFGLTANSPDTHLLLEGOSLTLTLESPPGSSPSVQCRSPRGKNIQGG 60

DB 167 KTLVSQLELDSDGTWTCTVLOK 191
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      61 RTLSVPLERQDSGTWTCTVNSODK 85

RESULT 20
077601 PRELIMINARY; PRT; 86 AA.
ID 077601;
AC 077601;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE (Fragment).
OS Lophocobus albigena albigena.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Lophocobus.
OC NCBI_TaxID=75566;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris E.E., Disocell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabey (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR HSSP; AF057391; AAC25135.1; -.
DR HSSP; P01730; LCDY.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR InterPro; IPR007110; Ig-like.
KW Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT NON_TER 1
FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 49 78 BY SIMILARITY.
FT NON_TER 86
SQ SEQUENCE 86 AA; 9463 MW; 2BD97A88464FB9AB CRC64;

Query Match      16.7%; Score 388; DB 6; Length 86;
Best Local Similarity 87.1%; Pred. No. 1.9e-22;
Matches 74; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 107 YICEVEDQKEEVOLVFGLTANSPDTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGG 166
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      1 YICEVEDKKEVELLVFGLTANSPDTHLLLEGOSLTLTLESPPGSSPSVQCRSPRGKNIQGG 60

DB 167 KTLVSQLELDSDGTWTCTVLOK 191
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      61 RTLSVPLERQDSGTWTCTVNSODK 85

RESULT 21
077600 PRELIMINARY; PRT; 86 AA.
ID 077600;
AC 077600;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)

```

```

DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
DE (Fragment).
OS Lophocobus aterrimus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Lophocobus.
OC NCBI_TaxID=75566;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris E.E., Disocell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabey (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR HSSP; AF057390; AAC25134.1; -.
DR HSSP; P01730; LCDY.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR InterPro; IPR007110; Ig-like.
KW Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.
FT NON_TER 1
FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT CARBOHYD 79 79 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT DISULFID 49 78 BY SIMILARITY.
FT NON_TER 86
SQ SEQUENCE 86 AA; 9476 MW; A9D97A88464FB9AB CRC64;

Query Match      16.5%; Score 383; DB 6; Length 86;
Best Local Similarity 85.9%; Pred. No. 4.7e-22;
Matches 73; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 107 YICEVEDQKEEVOLVFGLTANSPDTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGG 166
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      1 YICEVEDKKEVELLVFGLTANSPDTHLLLEGOSLTLTLESPPGSSPSVQCRSPRGKNIQGG 60

DB 167 KTLVSQLELDSDGTWTCTVLOK 191
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
      61 RTLSVPLERQDSGTWTCTVNSODK 85

RESULT 22
013969 PRELIMINARY; PRT; 71 AA.
ID 013969;
AC 013969;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-JAN-1999 (Tremblrel. 09, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE CD4 protein (Fragment).
DB CD4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zverev V.V., Blinov V.M., Nedospasov S.A.;
RT "Splice-mediated insertion of antisense and sense Alu repeats in human
RT CD4 gene: identification of three exons of CD4 mRNA.";
RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Zverev V.V., Sidarov A.V., Nedospasov S.A., Maliushova V.V.,
RA Udolova I.A., Andhaparkide O.G., Blinov V.M.;
RT "Nucleotide sequence of two exons of the human T-lymphocyte CD4

```

RT receptor gene).";  
 RL Vopr. Virusol. 40:100-102 (1995).  
 DR EMBL: X87579; CAA60883.1; -.  
 DR EMBL: S79267; AAB35273.1; -.  
 DR PIR: I60082; I60082.  
 DR HSSP: P01730; 1CDY.  
 DR GO: GO:0004872; F:receptor activity; IEA.  
 DR InterPro: IPR007110; IG-like.  
 DR PROSITE: PS50835; IG\_LIKE; 1.  
 KM Receptor.  
 FT NON\_TER 71  
 SQ SEQUENCE 71 AA; 7844 MW; ASC9D84816135C86 CRC64;

Query Match 15.4%; Score 357; DB 4; Length 71;  
 Best Local Similarity 98.6%; Pred. No. 3.6e-20;  
 Matches 70; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVPFRLLVLQALLPAAATQGNKRVYLGKKGDVVELTCTASQKKSIOFHMKNSNOIK 60  
 DB 1 MNRGVPFRLLVLQALLPAAATQGNKRVYLGKKGDVVELTCTASQKKSIOFHMKNSNOIK 60  
 QY 61 ILGNQSSFLTK 71  
 DB 61 ILGNQSSFLTK 71

## RESULT 23

Q29027 PRELIMINARY; PRT; 99 AA.

ID Q29027  
 AC 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE CD4, allele 1 (Fragment).  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_Taxid=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=#1183; TISSUE=Blood;  
 RX MEDLINE=93329116; PubMed=8335933;  
 RA Gustafsson K., Germana S., Sundt T.M., Sachs D.H., Leguern C.;  
 RT "Extensive allelic polymorphism in an exposed region of the  
 RT miniature. . . . .";  
 RL J. Immunol. 151:1365-1370 (1993).  
 DR EMBL: X65629; CAA46583.1; -.  
 DR PIR: I47131; S21461.  
 DR HSSP: P01730; 1CDY.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003596; IG\_v.  
 DR Pfam: PF00047; Ig; 1.  
 DR SMART: SM00406; IGV; 1.  
 DR PROSITE: PS50835; IG\_LIKE; 1.  
 FT NON\_TER 1  
 FT NON\_TER 99  
 SQ SEQUENCE 99 AA; 11170 MW; 40BF080692CF5DOC CRC64;

Query Match 13.1%; Score 305.5; DB 6; Length 99;  
 Best Local Similarity 60.2%; Pred. No. 5.1e-16;  
 Matches 59; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

QY 32 KKGDVVELTCTASQKKSIOFHMKNSNOIKILGNQSSFLTKGPSKILNDRADRSRLMDPG 90  
 DB 1 KAGDLELPHCHSQQKKNLPFNWKNSTKTLGGHGSFWHTASTVELTSLRSLDKKNMWDHG 60

QY 91 NFPLIIKNLEKIEDSDTYICEVEDQKEEVQLVFGTLTAN 128  
 DB 61 SFPLIIKNLEVTDSGYICEVEDKRIEVLQVLRILTAS 98

## RESULT 24

Q29028

ID Q29028 PRELIMINARY; PRT; 99 AA.

AC Q29028  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE CD4, allele 2 (Fragment).  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_Taxid=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=#1183; TISSUE=Blood;  
 RX MEDLINE=93329116; PubMed=8335933;  
 RA Gustafsson K., Germana S., Sundt T.M., Sachs D.H., Leguern C.;  
 RT "Extensive allelic polymorphism in an exposed region of the  
 RT miniature. . . . .";  
 RL J. Immunol. 151:1365-1370 (1993).  
 DR EMBL: X65630; CAA46584.1; -.  
 DR PIR: I47132; S21462.  
 DR HSSP: P01730; 1CDY.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003596; IG\_v.  
 DR Pfam: PF00047; IGV; 1.  
 DR SMART: SM00406; IGV; 1.  
 DR PROSITE: PS50835; IG\_LIKE; 1.  
 FT NON\_TER 1  
 FT NON\_TER 99  
 SQ SEQUENCE 99 AA; 11390 MW; C2295BEA228318F CRC64;

Query Match 12.1%; Score 280.5; DB 6; Length 99;  
 Best Local Similarity 56.1%; Pred. No. 4.3e-14;  
 Matches 55; Conservative 19; Mismatches 23; Indels 1; Gaps 1;

QY 32 KKGDVVELTCTASQKKSIOFHMKNSNOIKILGNQSSFLTKGPSKILNDRADRSRLMDPG 90  
 DB 1 KAGDLELPHCHSQQKKNLPFNWKNSTKTLKSHNNLWPKASVTELSRLSKKNMWDHG 60

QY 91 NFPLIIKNLEKIEDSDTYICEVEDQKEEVQLVFGTLTAN 128  
 DB 61 SFPLIIKNLEVTDSGYICEVEDKRIEVLQVLRILTAS 98

## RESULT 25

O88650 PRELIMINARY; PRT; 120 AA.

ID O88650  
 AC O88650  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE T-cell surface glycoprotein CD4 (Fragment).  
 GN CD4.  
 OS Marmota monax (Woodchuck).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;  
 OC Marmota.  
 OX NCBI\_Taxid=9995;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Periphereal blood;  
 RX MEDLINE=20094937; PubMed=10627561;  
 RA Guo J.T., Zhou H., Liu C., Aldrich C., Saputelli J., Whitaker T.,  
 RA Barrera M.I., Mason W.S., Seeger C.;  
 RT "Apoptosis and regeneration of hepatocytes during recovery from  
 RT transient hepatitis virus infections.";  
 RL J. Virol. 74:1495-1505 (2000).  
 DR EMBL: AF082497; AAC32621.1; -.  
 DR HSSP: P01730; 1WIO.  
 DR GO: GO:0016020; C:membrane; IEA.  
 DR GO: GO:0006955; P:immune response; IEA.  
 DR InterPro: IPR000973; CD4\_TcRg.  
 DR InterPro: IPR007110; IG-like.  
 DR PRINTS: PR00692; CD4TCANTIGEN.

```
FT NON TER 1 1
FT NON TER 120 120
SQ SEQUENCE 120 AA; 13623 MW; 44967B512D520195 CRC64;

Query Match 12.0%; Score 278.5; DB 11; Length 120;
Best Local Similarity 48.3%; Pred. No. 8e-14;
Matches 71; Conservative 9; Mismatches 38; Indels 29; Gaps 3;

QY 308 LHOENVLVNMRATQLOKNCILCEVWGPTSPKLMSTLKENKEAVSKREKRVWMLNPAQM 367
   1 LHOENVLVNMRATQLOKNCILCEVWGPTSPKLMSTLKENKEAVSKREKRIKRVNPPAGM 60
   1 LHOENVLVNMRATQLOKNCILCEVWGPTSPKLMSTLKENKEAVSKREKRIKRVNPPAGM 60

DB 368 WOCILSDSGOVLLSENIKVLPTWSTPVNPPASALPAPPTGSALPDPTASALPDPPASA 427
   61 WOCILRGRGKVLVDFOQDVEPTTELNOQPMFLAV---IIGGAL----- 100

QY 428 LPAALAVISFLGLGLGV-ACVLARTR 453
   101 -----SFLLAGLGLCFCCVACRHR 119

RESULT 26
Q90WB5 PRELIMINARY; PRT; 482 AA.
AC Q90WB5;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE 1-cell surface glycoprotein CD4 precursor.
OS Anas platyrhynchos (Domestic duck).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anas.
OX NCBI_TaxID=8639;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Chan S.W.S., Middleton D.L., Lundqvist M., Marr G.W., Higgins D.A.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF378701; AAK59279.1; -.
DR InterPro: IPR003599; IG.
DR InterPro: IPR007110; IG-like.
DR Pfam: PF00047; IG; 4.
DR SMART: SM00409; IG; 3.
DR PROSITE: PSS0835; IG_LIKE; 2.
DR Signal.
KW SIGNAL.
FT SIGNAL.
SQ SEQUENCE 482 AA; 54624 MW; BDA0E8EC6172AD0B CRC64;

Query Match 11.7%; Score 271.5; DB 13; Length 482;
Best Local Similarity 25.0%; Pred. No. 1.9e-12;
Matches 111; Conservative 69; Mismatches 169; Indels 95; Gaps 19;

QY 8 RHLLVLQALLPATQGNKVLGKGDVTELTCTASQKSIQPHWK-----NSNQIK 60
   12 RAVFVLLQGLL--THIMHQOQIGVEGEVILNC---KKHKDQVTKWEYDAGSSAIIIG 66
   61 ILNGSGFLTKGSKLANDRSDRSRLMDQGNPPLIINKLIEDSPYICEVEDQKEEVOL 120
   67 ILGK--IFKGAAPMSDRSET-----NNSKHLKVNLRISDAGTYICGSDRNSISL 118
   121 LVFGLTANSPTHLQOGSLTLT--LESPGSSPS-----VOCRRPRGNK 162
   119 HAVKLTISNGYFLPGDDELITVHKSPKQPRPSITLPHSHSRVTPEVLQNETPO--- 175
   163 IQGKTLVSQLELDQSGTWTCTVLQNKVEFKI--DIVLAFOKAS--SIYVKEGEQV 219
   176 ---KYALKVQLOPDSGTWICNNHSDSPSINENISFNKVLGPEKTLERMTAAVDSTV 232
   220 EFSFPLATVEKLTG-----SGELMQAERASSKSWITFDLKNKEVSVKRTQDPKQM 274
   233 TLSEWHLNF---RIKGWKEPFTQGLMQEGNA-----ITYELDFNATADGELRETK-- 280
   275 GKSLPLHLTLPLQALP-----QYAGSGNLTALAEKTKGLHGVNVLVVMRATQ-- 321
```

```
DB 281 -KKSQALTEIPENKRDSTVEVKIHKIQLKHSGETQGLLYNRRIYQSKTELVMVQNSAP 339
   322 -----LQKNTLCEWGPSPKLM-----SLKENKEAKVSKREKRVWMLNPAQM 368
   340 PGPPLKGAEMTLLCQVSSPPLPVNHLMERVNGTKMDGKSKOS--ETKVEKVTAVGM 397
   369 QCLSDSGOVLLSENI--IKVLPW 390
   398 NCHLMEDNNKLSLNTYVEAPTW 421

RESULT 27
Q9MEV7 PRELIMINARY; PRT; 487 AA.
AC Q9MEV7;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE CD4 protein precursor.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H.B2; TISSUE=Thymus;
RX MEDLINE=99218434; PubMed=10201936;
RA Koskenen R., Lammiuaki U., Tregaskes C.A., Salomonsen J., Young J.R.,
RA Vainio O.;
RT "Cloning and modeling of the first nonmammalian CD4.";
RT J. Immunol. 162:4115-4121 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=B12;
RA Koskenen R.H., Salomonsen J., Tregaskes C.A., Young J.R.,
RA Goodchild M., Bumstead N., Vainio O.;
RT "The chicken CD4 gene has remained conserved in evolution.";
RL Immunogenetics 0:0-0 (2002).
DR EMBL: Y12012; CAA72740.1; -.
DR EMBL: AJ401223; CAC82027.1; -.
DR InterPro: IPR007110; IG-like.
DR Pfam: PF00047; IG; 3.
DR PROSITE: PSS0835; IG_LIKE; 2.
DR Signal.
KW SIGNAL.
FT SIGNAL.
SQ SEQUENCE 487 AA; 54985 MW; B9CBA92EC9F7F45B CRC64;

Query Match 10.5%; Score 245; DB 13; Length 487;
Best Local Similarity 23.9%; Pred. No. 2.1e-10;
Matches 121; Conservative 66; Mismatches 193; Indels 126; Gaps 22;

QY 10 LLLVLQALLPATQGNKVLGKGDVTELTCTA--SQKSIQPHWK-----SNOIKIG 63
   14 VILVLQGLTPMAQEQEI--GIAGKEVILSCALINQDQGTCTMKYKKEVSTIIS 72
   64 NQGSFLTKGSPKLANDRSDRSRLMDQGNPPLIINKLIEDSPYICEVEDQKEEVOLVF 123
   73 KQVFGKAPM-----THRSGLNSSKKLKVSLSLDAGIYTCACVSPVVISLHVF 125
   124 GLTANSPTHLQOGSLTLT--ESPPSSPS-----VOCRRPRGNK 165
   126 KLTISNGHFLTNMEDLELTLMQNSHSHQPHLSIKLFNINNDIVTEILQEBAPQ--KYI-- 182
   166 GKTLVSQLELDQSGTWTCTVLQNKVEFKI--DIVLAFOKAS--SIYVKEGEQV 222
   183 ---LKKQKALDSGTMCHVSNPSINQNISFDVKVIGFEKERLEIITYTVGNATIS 239
   223 FPLATVEKLTG--SGELMQAERASSKSWITFDLKNKEVSVKRTQDPKQMGKPL 280
   240 WRLNFKIKWKEGFTKLMWBPQNTAIEHLNFS-----VTTHQELHKTCK--SN 288
```





```

DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR003989; VCAM-1.
DR Pfam; PF00047; Ig; 5.
DR PRINTS; PR01472; ICAMVCAM1.
DR PRINTS; PR01474; VCAM1.
DR SMART; SM00408; IGC2; 4.
DR PROSITE; PS50835; IG_LIKE; 5.
DR Immunoglobulin domain; Signal.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 739 VASCULAR CELL ADHESION MOLECULE 1.
SQ SEQUENCE 739 AA; 81136 MW; A5AAD1172F7FB96 CRC64;

Query Match
Best Local Similarity 22.0%; Pred. No. 0.00089;
Matches 86; Conservative 60; Mismatches 155; Indels 90; Gaps 14;

QY 33 KQDTVELTCTASQKSIQFHWNKSNQIKLGNQSFLLTKGPSKLNDRADSRSLMDQGNF 92
DB 238 EGAAYVTMTCASEGLPAPPEIFWMSKK-----LDNGVQLQL-----SGNA 274
QY 93 PLTIKNIKIEDSDTYICE---VEDQKEVQLLV-----FGLTNSDTHLLQGSLLTLT 142
DB 275 TLTLIIMRMEDSGIYVCEGNVLVGRDKTEVELIVQEKPTVIDISPGSVAAQVGDVSLT 334
QY 143 LESPPGSSPSVQCRSPRGKNIQ-----GKTLVSQLELQDSGTWTCTVLQNOKVBF 195
DB 335 CAAVGCDSPFSFWRQTQDSPNGEVNRDEGATSTLTISPVGVEDEHSYLTCTVTCQRKLEK 394
QY 196 KIDIVVLAQKASIIYKKEGEVERSFPLA-----FTVEKLTGSG 236
DB 395 TLQVEYVSF-----PEDPEIEISGLVHGRPTVNTCTVNPVYFPDHLIELLKGET 445
QY 237 ELWMOAERAS-SSKSWITTFDLKNKEYSVKRVOTDPFLQMGKCL---PLHLTLPLALP-Q 290
DB 446 TLANKLREIGTGS---LETKLEMTFTPTAED---TGKALVCLAKLHSSQMESEPPQ 498
QY 291 YAGSGNLTALAEAKTKLHQEVNLVVMRATQLOKNTLCEWGPPTSFKMLSLKLENKAK 350
DB 499 RQGTQTLVYVAVAPKEPTIWPSPVPPEGSPV--NLTCSDEGPTPTKILMSRQLKNGELQ 556
QY 351 VSKREKPVWLNPEAGMOCCLSDSQVLLF 381
DB 557 PLSQ-----NTTLSFMATKMGDSGIYVCE 580

RESULT 33
Q29123 PRELIMINARY; PRT; 538 AA.
ID Q29123;
AC Q29123;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Vascular cell adhesion molecule.
GN VCAM.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sub.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Aorta;
RX MEDLINE=94271236; PubMed=7516159;
RA Teang Y.T., Haekard D.O., Robinson M.K.;
RT "Cloning and expression kinetics of porcine vascular cell adhesion
molecule.";
RL Biochem. Biophys. Res. Commun. 201:805-812(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Aorta;
RX MEDLINE=96106491; PubMed=8525525;
RA Mueller J.P., Evans M.J., Cotfeli R., Rother R.P., Mattis L.A.,
RA Elliott E.A.;
RT "Porcine vascular cell adhesion molecule (VCAM) mediates endothelial

```

```

RT cell adhesion to human T cells. Development of blocking antibodies
RT specific for porcine VCAM.";
RL Transplantation 60:1299-1306(1995).
DR EMBL; LA3124; AAB59281.1; -.
DR HSBP; P19320; IYCA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0016337; P:cell-cell adhesion; IEA.
DR InterPro; IPR003987; ICAM VCAM-1.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR003989; VCAM-1.
DR Pfam; PF00047; Ig; 5.
DR PRINTS; PR01472; ICAMVCAM1.
DR PRINTS; PR01474; VCAM1.
DR SMART; SM00408; IGC2; 3.
DR PROSITE; PS50835; IG_LIKE; 4.
DR Immunoglobulin domain.
SQ SEQUENCE 538 AA; 58713 MW; 8A7CD36DDA2F0717 CRC64;

Query Match
Best Local Similarity 23.7%; Pred. No. 0.00063;
Matches 66; Conservative 49; Mismatches 90; Indels 73; Gaps 11;

QY 32 KKGDTVELTCTASQKSIQFHWNKSNQIKLGNQSFLLTKGPSKLNDRADSRSLMDQGN 91
DB 235 QEGDSVMMTCTSEGLPAPQISM-----SKLIDNGDQLL---SGN 271
QY 92 PFLTIKNIKIEDSDTYICE---VEDQKEVQLLV-----FGLTNSDTHLLQGSLLTL 141
DB 272 ATLTLIIMRMEDSGIYVCEGNVPVGTNRKEVELIVQAVPRDTTISVNSSTLEBSSVNM 331
QY 142 TLESPPGSSPSV---QCRSPRGKNIQGKTLVSQLELQDSGTWTCTVLQ---NOKV 193
DB 332 TCSSGFPAPKILMKSLRDGRLBPLSENTTLTLTSTMEBSGIYVCGINGNRRKEV 391
QY 194 EFKI-----DIVTLAFOKAS-----SIYKKE---GEVERSFPLA 226
DB 392 ELIIQAPKQDLIILFPSESVYEGPTVIISCTCGNVPPTLIILKKAEFTGDTVLKSTDA 451
QY 227 FTVEKL-----TGSGLWMOAERASSKSWITTFDLKNKE 260
DB 452 YTHRAPRLADGAVYECESKNEITGLQLS-ITLDVAGRE 488

RESULT 34
Q8NFP4 PRELIMINARY; PRT; 955 AA.
ID Q8NFP4;
AC Q8NFP4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glycosyl-phosphatidylinositol-MW.
GN GPI-M.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22077705; PubMed=12082541;
RA De Juan C., Iniesta P., Gonzalez-Quevedo R., Moran A.,
RA Sanchez-Penaute A., Torres A.J., Balibrea J.L., Diaz-Rubio E.,
RA Cruces J., Benito M.;
RT "Genomic organization of a novel glycosylphosphatidylinositol MAM gene
expressed in human tissues and tumors.";
RL Oncogene 21:3089-3094(2002).
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
DR EMBL; AF478693; AAM77220.1; -.
DR Genew; HGNC:19267; MDGAL.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0016163; F:nitrogenase activity; IEA.
DR GO; GO:0009399; P:nitrogen fixation; IEA.
DR InterPro; IPR008957; FN_III-like.

```



DR InterPro: IPR003599; IG.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003598; IG c2.  
 DR InterPro: IPR000998; MAM\_domain.  
 DR InterPro: IPR000318; Nitrogene\_comp1.  
 DR Pfam: PF00047; IG; 6.  
 DR Pfam: PF00629; MAM; 1.  
 DR PRINTS: PRO0020; MAMDOMAIN.  
 DR SMART: SM00409; IG; 6.  
 DR SMART: SM00408; IGc2; 6.  
 DR SMART: SM00137; MAM; 1.  
 DR PROSITE: PSS0835; IG LIKE; 6.  
 DR PROSITE: PSS0060; MAM 2; 1.  
 DR PROSITE: PSS0699; NITROGENASE 1.1; 1.  
 DR GlycoProfile: Immunoglobulin domain.  
 KM SEQUENCE 955 AA; 105790 MW; BD41A1EB10A05962 CRC64;  
 SQ  
 Query Match 6.8%; Score 157; DB 4; Length 955;  
 Best Local Similarity 23.3%; Pred. No. 0.0031;  
 Matches 112; Conservative 64; Mismatches 181; Indels 124; Gaps 25;

QY 36 TVELTCTASQKSKIQFHWKNSNQIKILGNQSGFLTKGPSKLNDRADSRSLMDQGNFPLI 95  
 DB 152 TVFLRCTVNSNPAPFIMKRGSD-----TLSHSQDNG-VDIYEPLTYQGETKVL 199  
 QY 96 -IKNLKIEDSDTYICEVDQKE---EYQLVFGITANSDFHILQGGSLTILTESPPGSSP 151  
 DB 200 KLNKLRPODYASTYCVQSVNVCIGIPDKAITFLTNTTAPPAK-LSVNETLLVNPGENV 258  
 QY 152 SVQC-----RSPRGKNIQGGKTLVSQLELDQSGTWTCTVLQ----- 189  
 DB 259 TVGCLLTGSDPLPQLQWHSRGPLPLGALAQG-TLSPVQARDSDGYVCTATNNVGNP 317  
 QY 190 QKKVEFKIDIVLAFQASSIVK--KEGQVEFSPL-----AFVTEKLTGSGELW 229  
 DB 318 AKKT--VNLVLSMKATFQITPDVISEENIQGLQDGLKLSCHVDVAPQEKTYQ---W 371  
 QY 240 MQ-AERASSSKSWITPLNKKVESVKRVITQDPKIQMKPLH-----TLPGA- 287  
 DB 372 FKNGKPRMSKRL--VTRNDPELPAVTSLEL-----IDLHSDGYTLVCMASFPGAR 423  
 QY 288 LPOYASGNTLLEAKTGKLGHEVNLVVRATOLQKLTCEVWGPTSPKMLSLKLENK 347  
 DB 424 VPDLSVEVNISSEVPTTISVPRKRAVTVREGS-PAELQCEVRGKRPVILWS--RVDK 480  
 QY 348 EAK----VSKREKPVWLNP-----AGMOC-----LSDSGVLE----- 381  
 DB 481 EALLPFGGLPLEETPDGKRLERVSRDMSGTYRCQTARVNGFVVRPREAQVQINVQFPPE 540  
 QY 382 ---SNIVLPTWSTPVPVPRASALPAP-----TGSALPDPOPTASALPPPPASAL 428  
 DB 541 VEPDSQDVRQALGRVLLRCSLLRGSPQIASAVMRFKQLLPP-----PVPPAAAEA 594  
 QY 429 P 429  
 DB 595 P 595

RESULT 35  
 ID Q28939 PRELIMINARY; PRT; 538 AA.  
 AC Q28939;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Vascular cell adhesion molecule precursor.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RC TISSUE=Aorta;  
 RP

RX MEDLINE=94271236; PubMed=7516159;  
 RA Tsang Y.T., Haekard D.O., Robinson M.K.;  
 RT "Cloning and expression kinetics of porcine vascular cell adhesion  
 molecule.";  
 RL Biochem. Biophys. Res. Commun. 201:805-812(1994).  
 DR EMBL: U08351; AAA21542.1; -.  
 DR PIR: JC2457; JC2457.  
 DR HSSP: P19320; 1VCA.  
 DR GO: GO:0016020; C:membrane; IEA.  
 DR GO: GO:0016537; P:cell-cell adhesion; IEA.  
 DR InterPro: IPR003987; ICAM VCAM-1.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003598; IG c2.  
 DR InterPro: IPR003989; VCAM-1.  
 DR Pfam: PF00047; IG; 5.  
 DR PRINTS: PRO1472; ICAMVCAM1.  
 DR PRINTS: PRO1474; VCAM1.  
 DR SMART: SM00408; IGc2; 3.  
 DR PROSITE: PSS0835; IG LIKE; 4.  
 KM Immunoglobulin domain; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 538 VASCULAR CELL ADHESION MOLECULE.  
 SQ SEQUENCE 538 AA; 58795 MW; EC29D1B224F7261 CRC64;  
 Query Match 6.7%; Score 155.5; DB 6; Length 538;  
 Best Local Similarity 23.4%; Pred. No. 0.0018;  
 Matches 65; Conservative 49; Mismatches 91; Indels 73; Gaps 11;

QY 32 KKGDVLELTCTASQKSKIQFHWKNSNQIKILGNQSGFLTKGPSKLNDRADSRSLMDQGN 91  
 DB 235 QEDSMWMTCTSGPLAPISW-----SKLDNDQQLL-----SGN 271  
 QY 92 FPLINKLKIEDSDTYICE---VEDQKEVQLLV-----FGLTANSDFHILQGGSLT 141  
 DB 272 ATLTLTLMWMBDSGIYCEGVNPGVGNRKREVELTVVAPRDTTISVNPSTLEEGSSVM 331  
 QY 142 TLESPPGSSPSV-----QCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLQ-----NOKKY 193  
 DB 332 TCSSDGFPAFKIMSKKLDGNIPLSENTTLTLTSTKEDSDIYCEGINQAGIRKEY 351  
 QY 194 EFKI-----DIVLAFQAS-----SIYKKE---GEQVEFSFPLA 226  
 DB 392 ELIIQAPDLOQTAPSPSEVKEGTVIISCTGVNPPILLIKKAEAGDFVLKSTDGA 451  
 QY 227 FTYEKL-----TSGGLMQAERASSSKSWITPLDKKE 260  
 DB 452 YTIHRLADAGVCECSKNEIGLQLRS-ITLDVKGRE 488

RESULT 36  
 ID Q90DES PRELIMINARY; PRT; 31 AA.  
 AC Q90DES;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE CD4-55 kDa glycoprotein (Fragment).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP MEDLINE=92072595; PubMed=1961196;  
 RX Lederman S., Demartino J.A., Daugherty B.L., Foeldvari I.,  
 RA Yellin M.J., Cleary A.M., Berkowitz N., Lowy I., Braunstein N.S.,  
 RA Mark G.B.;  
 RT "A single amino acid substitution in a common African allele of the  
 RT CD4 molecule ablates binding of the monoclonal antibody, OKT4.";  
 RL Mol. Immunol. 28:1171-1181(1991).  
 DR HSSP: P01730; 1WIO.  
 DR InterPro: IPR007110; IG-like.  
 FT NON\_TER 1 1

```

FT  NON TER      31      31
SQ  SEQUENCE      31 AA; 3658 MW;  A7E9C61F5DPCFCEFCRC64;

Query Match
Best Local Similarity  6.5%; Score 152; DB 4; Length 31;
Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy      250 SWITFDLKNKEVSVKRVTDOPKLOMGKRLPL 280
      |||||
Db      1 SWITFDLKNKEVSVKRVTDOPKLOMGKRLPL 31

RESULT 37
Q92626 PRELIMINARY; PRT; 1496 AA.
AC  Q92626;
DT  01-FEB-1997 (TrEMBLrel. 02, Created)
DT  01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  MVEL0BLAST KIAA0230 (Fragment).
GN  KIAA0230.
OS  Homo sapiens (human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Bone marrow;
RX  MEDLINE=97191544; PubMed=9039502;
RA  Nagae T., Seki N., Ishikawa K., Ohira M., Kawabayashi Y., Ohara O.,
RA  Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT  "Prediction of the coding sequences of unidentified human genes. VI.
RT  The coding sequences of 80 new genes (KIA0201-KIA0280) deduced by
RT  analysis of cDNA clones from cell line KG-1 and brain.";
RL  DNA Res. 3:321-329(1996).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Melanoma;
RX  MEDLINE=95048383; PubMed=7959781;
RA  Weiler S.R., Taylor S.M., Deans R.J., Kan-Mitchell J., Mitchell M.S.,
RA  Trent J.M.;
RT  "Assignment of a human melanoma associated gene MG50 (D2S448) to
RT  chromosome 2p25.3 by fluorescence in situ hybridization.";
RL  Genomics 22:243-244(1994).
RN  [3]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Melanoma;
RA  Mitchell M.S., Kan-Mitchell J., Minev B., Edman C., Deans R.J.;
RT  "Identification of a novel melanoma gene (MG50) - likely the gene for
RT  IL-1 receptor antagonist - which encodes epitopes recognized by human
RT  cytolytic T lymphocytes.";
RL  Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR  EMBL; D86983; BA013219.1; -.
DR  EMBL; AF200348; AA06354.1; -.
DR  HSP; P05164; IXP.
DR  GO; GO:0005152; F:interleukin-1 receptor antagonist activity; NAS.
DR  GO; GO:0006955; P:immune response; NAS.
DR  InterPro; IPR002007; Anim_peroxidase.
DR  InterPro; IPR007110; Ig-like.
DR  InterPro; IPR003598; Ig_c2.
DR  InterPro; IPR001611; LRR.
DR  InterPro; IPR000483; LRR_Cterm.
DR  InterPro; IPR000372; LRR_Nterm.
DR  InterPro; IPR003591; LRR_Typ.
DR  InterPro; IPR002016; Peroxidase.
DR  InterPro; IPR001007; VWF_C.
DR  Pfam; PF03098; An_peroxidase; 1.
DR  Pfam; PF00047; Ig; 4.
DR  Pfam; PF00560; LRR; 5.
DR  Pfam; PF01463; LRRCT; 1.
DR  Pfam; PF00093; vwc; 1.
DR  PRINTS; PR00457; ANPEROXIDASE.
DR  SMART; SM00408; IGC2; 4.

```

```

DR  SMART; SM00082; LRRCT; 1.
DR  SMART; SM00013; LRNT; 1.
DR  SMART; SM00369; LRR_Typ; 4.
DR  SMART; SM00214; VWC; 1.
DR  PROSITE; PS00835; IG_LIKE; 4.
DR  PROSITE; PS02992; PEROXIDASE_3; 1.
DR  PROSITE; PS01208; VWF_C_1; 1.
DR  PROSITE; PS0184; VWF_C_2; 1.
DR  Immunoglobulin domain.
FT  NON TER      1
SQ  SEQUENCE      1 AA; 167209 MW;  E9B9A7069BF1ABEF CRC64;

Query Match
Best Local Similarity  21.5%; Score 152; DB 4; Length 1496;
Matches 100; Conservative 57; Mismatches 155; Indels 154; Gaps 19;

Qy      34 GDTVELCTASQKSIQFHMKNSNOIKILGNQGSFLTTPSKLNDPADSRRLMOGNEP 93
      |||||
Db      277 GNTVVFYTCRAEENPKREIIMLRNN-----NELSMKTDTSRLNLDDGT-- 318

Qy      94 LIHKNLKIEDSDTYICEV-----EDQKEVOLLVFGLTAN-----SPTHLQGSQSLTL 141
      |||||
Db      319 LMTQNTQETDQCIYQMAKNVAGEVKTQGVTLRYGSPARPFVYIQPONTVELVESVTL 378

Qy      142 ----TLSEPPGSSPSVQCRSP-----RCNIOGKTLVSQLELDQSGTWCTVLQNKVY 193
      |||||
Db      379 ECSATGHPPEPRISWRGRDRLPLVPDPRVNITPSGGLYIQNVQDPSGEYACSATNNIDSV 438

Qy      194 EKKIDIVLAFQKASIIYKK---EGEYVPSFLAFYVEKLTGSGELMQWERASSSSS 250
      |||||
Db      439 HATFATIIYQALPEQFVTVPDRVIEGQTVDF-----QCEAKGNPP 479

Qy      251 WITFDLKNKEVSVK---VTQPKLOM-----GKULPLHLTL-PQ 286
      |||||
Db      480 VIATWKGSQSLSDVRHLVSSGTLRISGVALHDQGYECQAVNIGSKVAHLTVQPR 539

Qy      287 ALPOYAGSGNLTALAEATGKLGQEVNLVMEATQLQKNLTGCEVWGPTSPKMLSLKLEN 346
      |||||
Db      540 VTPVFPASIPSDTTV-----EVGANV-----QLPSSGSGEPFATW-----N 576

Qy      347 KEAKSKKEKPVWVUNPE-----AGMOCGL-----SSGQVLTESNKKVLP 389
      |||||
Db      577 KQG-VQVTESGKFHISPEGFLLTINDVGPADARVYCVARNTIGSASVSVLSVNV----- 630

Qy      390 WSTPVHPRASALPAPPTGSALPDPTASALPPPPASALPALAVY 435
      |||||
Db      631 -----PD---VSRNGDPVATISIVEAIAIV 652

RESULT 38
Q865F2 PRELIMINARY; PRT; 739 AA.
AC  Q865F2;
DT  01-JUN-2003 (TrEMBLrel. 24, Created)
DT  01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  Adhesion molecule VCAM-1.
GN  VCAM-1.
OS  Oryctolagus cuniculus (Rabbit).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX  NCBI_TaxID=9986;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  Schnurr K., Banning A., Kupper D., Muller-Schmehl K.,
RA  Brigelius-Flohe R.;
RT  "Modulation of basal and interleukin-1-induced adhesion molecule
RT  expression by phospholipid hydroperoxide glutathione peroxidase and
RT  15-lipoxygenase in rabbit aortic smooth muscle cells.";
RL  Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AT212510; AA052742.1; -.
DR  GO; GO:0016020; C:membrane; IEA.
DR  GO; GO:0016337; P:cell-cell adhesion; IEA.

```

```
DR InterPro: IPR003987; ICAM_VCAM-1.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR007110; Ig_1ike.
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003989; VCAM-1.
DR Pfam: PF00047; Ig; 5.
DR PRINTS: PRO1472; ICAMVCAM1.
DR PRINTS: PRO1474; VCAM1.
DR SMART: SM00409; Ig; 5.
DR SMART: SM00408; IgC2; 5.
DR PROSITE: PS50835; IG_LIKE; 5.
SQ SEQUENCE 739 AA; 81806 MW; 7AD0D32511E19342 CRC64;

Query Match 6.5%; Score 151.5; DB 6; Length 739;
Best Local Similarity 19.9%; Pred. No. 0.0057;
Matches 83; Conservative 71; Mismatches 149; Indels 115; Gaps 17;

QY 32 KKGDTVELTCTASQKSIQFHWKNSNQIKILNGSFLTKGPKLNDRADRSRLMDQGN 91
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 237 QEGGSVMTCTSSSEGLPYPPELFW-----SKQDNGNLQRL---SGN 273
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 92 FPLIKNLKIEDSDTYICEVEDQ---KEEVQLLVFGLTANSDTHLLQGSLTLLTESPP 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 274 ATLTLAMRMEDSGIYVCEGVNQIGKSRKEVELIV-----QEKPFVEISP 319
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 148 G-----SSSVQCRSPRGKNIQCKT-----LSVSQLELDQSG 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 GPRIAAQIGDPVVLTCVRGCEPFSFWRQIDSPILGQVTSKGTSLTLSPVSENEH 379
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 181 TWCTVVLQNGKVFYKIDIVLAFQKASIVYK---KEGEQVPS-----FPL-APTVE 230
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 380 SYLCTVYCGHKLEKGIQVELYSPRPDEIELSGPPNGRPVYSCVPVNYFPDKLIE 439
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 231 KLTGSGELMWQAEPASSKSMI--TFDLKKNKVSRYKVTQDPKIQ-MKKL----PLHL- 282
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 440 LKLG-----ETPMKNKEFLLEBEDKSLKTSLEMTFPIPTMEDTGKVLVCAKHLIDE 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 283 -----TLPOLPQYAGSNLTLALEAKTGKLNQEVNLVYMRATQLOKNTCEWGTSP 336
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 492 MEPEPKRQSTQPLFVNVAPRDIAVWVSPSSIVEEGRSV-----NMTCSSYGLPAP 542
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 337 KLMLSLKLENKEAKVSKREKPVWVLA--PEAGMWOC---LLSDSGVLLSNIKVLVP 388
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 543 KILMSRQLKNGDQLPLSENTTLALITKLEDSGIYVCEGINLAKSKVEVELYIQVAP 600
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 39
Q80U60 PRELIMINARY; PRT; 1431 AA.
AC Q80U60;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MxiA0230 protein (Fragment).
GN MxiA0230.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain.
RX MEDLINE=22579291; Pubmed=12693553;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Rec. 10:35-48(2003).
DR EMBL; AK122223; BAC65505.1; -.
DR GO; GO:0004601; F:peroxidase activity; IEA.
DR GO; GO:0006979; P:response to oxidative stress; IEA.
```

```
DR InterPro: IPR002007; Anim_peroxidase.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR007110; Ig_1ike.
DR InterPro: IPR003598; Ig_c2.
DR InterPro: IPR003596; Ig_v.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR000483; LRR_Cterm.
DR InterPro: IPR007092; LRR_SDS22.
DR InterPro: IPR003591; LRR_tyr.
DR InterPro: IPR002016; Peroxidase.
DR InterPro: IPR001007; VWF_C.
DR Pfam: PF03098; An_peroxidase; 1.
DR Pfam: PF00047; Ig; 4.
DR Pfam: PF00560; LRR; 5.
DR Pfam: PF00463; LRRCT; 1.
DR Pfam: PF00093; VWC; 1.
DR PRINTS: PRO0457; ANPEROXIDASE.
DR SMART: SM00409; Ig; 4.
DR SMART: SM00406; IgC2; 4.
DR SMART: SM00082; LRRCT; 1.
DR SMART: SM00366; LRR_PS; 4.
DR SMART: SM00369; LRR_TYP; 5.
DR SMART: SM00214; VWC; 1.
DR PROSITE: PS50835; IG_LIKE; 4.
DR PROSITE: PS50292; PEROXIDASE_3; 1.
DR PROSITE: PS01208; VWF_C_1; 1.
DR PROSITE: PS50184; VWF_C_2; 1.
FT NON TER
SQ SEQUENCE 1431 AA; 160591 MW; 6BA952436DA54B72 CRC64;

Query Match 6.2%; Score 145; DB 11; Length 1431;
Best Local Similarity 20.4%; Pred. No. 0.045;
Matches 96; Conservative 59; Mismatches 153; Indels 162; Gaps 20;

QY 34 GDTVELTCTASQKSIQFHWKNSNQIKILNGSFLTKGPKLNDRADRSRLMDQGNFP 93
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 213 GNTVYFTCAENKPKKEIIMLRN-----NELSMKTDRLNLTDDGT-- 254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 94 LIKLNKIEDSDTYICEV-----EDQKEVQLLVFGLTAN-----SDTHLLQGSLTL 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 255 LMIQNTQAEADGEGVQCMAKNVAGEAKTQEVTLRLYLSPARPFEVIQPTTEVLVGS--V 312
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 142 TLESPPGSSPSVQCRPRGK-----NIQGGTSLSVSQLELDQSGTWCTVLOKNG 191
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 313 TLECSATGHPLPQITWTRGDRTPPLIDPRVNTTPSGGLTYIQNVAQSDSEYTCFASNSVD 372
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 192 KVEFKIDIVLAFQK-----ASIVYKKEGEQVFSFPLAFVTEKLTGSGELMWQAEPA 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 373 SIHATAFITVQALPQTVTPQSRVLI--EGQYVF-----QCAAG 411
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 247 SSKSWITFDPLKKNKESVKR---VTQDPKIQM-----GKKPLHLIT 283
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 412 HPQPVIAWTKGSGQLSVDRHLVLSGTLRISGVALHDQGYEQCAVNIIGSKVVAHHT 471
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 284 L-PQLPQYAGSNLTLALEAKTGKLNQEVNLVYMRATQLOKNTCEWGTSPPKMLSL 342
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 472 VQRPVTPVPAASIPS-----DMTVEGVTVQ--LPCSSQGEPEPAITW-- 511
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 343 KLENKEAKVSKREKPVWVLA--PEAGMWOC---LLSDSGVLLSNIKVLVP 385
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 512 ---NKDG-VQVTESGGFHISPEGFLLINDVGNADAGRYECVARNITGYASVSNVLSVNV- 566
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 386 VLPFTWSTPVNPRASALPAPPTGSLPDPQTASALPDPAPASALPALAAVI 435
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 567 -----PD---VSRNGDPVATSIIVEAIATV 588
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 40
Q812P8 PRELIMINARY; PRT; 437 AA.
ID Q812P8;
AC Q812P8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
```



DR GO:0004197; F:cysteine-type endopeptidase activity; IEA.  
 DR GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR000152; Asx\_hydroxyl\_S.  
 DR InterPro: IPR000875; Cectropin.  
 DR InterPro: IPR001434; DUF11.  
 DR InterPro: IPR001881; EGF\_Ca.  
 DR InterPro: IPR006209; EGF\_like.  
 DR InterPro: IPR009017; GPF\_like.  
 DR InterPro: IPR007110; Ig-Like.  
 DR InterPro: IPR003598; Ig\_c2.  
 DR InterPro: IPR00169; SHprot\_acslike.  
 DR InterPro: IPR000884; TSP1.  
 DR InterPro: IPR002035; VWF\_A.  
 DR Pfam: PF00008; EGF\_5.  
 DR Pfam: PF00047; Ig; 44.  
 DR Pfam: PF00090; tsp\_1; 6.  
 DR SMART: SM00179; EGF\_CA; 7.  
 DR SMART: SM00408; IGC2; 43.  
 DR SMART: SM00209; TSP1; 6.  
 DR SMART: SM00327; VMA; 1.  
 DR TIGRfam: TIGR01451; Bant\_repeat; 9.  
 DR PROSITE: PS00010; ASX\_HYDROXYL; 5.  
 DR PROSITE: PS00268; CECTROPIN; 1.  
 DR PROSITE: PS01186; EGF\_2; 3.  
 DR PROSITE: PS01187; EGF\_CA; 8.  
 DR PROSITE: PS00639; THIOL\_PROTEASE\_HIS; 1.  
 DR PROSITE: PS50092; TSP1; 6.  
 DR EGF-like domain; Immunoglobulin domain.  
 KM SEQUENCE 5636 AA; 613660 MW; F000B319CED7B52C CRC64;  
 SQ

Query Match 6.2%; Score 143; DB 4; Length 5636;  
 Best Local Similarity 20.4%; Pred. No. 0.44; Indels 138; Gaps 21;  
 Matches 93; Conservative 56; Mismatches 168;

QY 34 GDTVELCTASQKSIQFHWKNSNQIKILNGSGFLTKGPKLNDRADSRSLMDQGNFP 93  
 DB 807 GSNVTLCTCYOQGYEPPIKRRLDNMPISFR--PFSVSSISQLRTGA----- 851  
 QY 94 LIINKLIEDSDTYICEVEDQ---KEEVOLLVFGLTA-----NSDTLLHQGSLTTL 143  
 DB 852 LFIINLWASDKGYICEAENQFGKIQGETTVTVGLVAPLIGISPSVANYIEGQQLTLP 911  
 QY 144 ESPPGS-----SPSVQCRSPRGKNIQSGKTLVSQLELDSGTWTC--- 164  
 DB 912 TLLAGNPIPERRWIKNSAMLLQNPYIVRS-----DGLHIERVOLDDGGGYTCVAS 963  
 QY 185 TVLQNKKEVEPKIDIVLAFQKASSIYVKKGEQVEFSPLAFTVEKLTGSGELMWAER 244  
 DB 964 NVAGTNNKTSVNVHVLPTIOHQGLISTEG--IPVTLF-----CKA 1004  
 QY 245 ASSSKSNITTDLKNKEVSVKRVTDPKLQNGKPLHLTLPLQALPYQVSGNLTALAEAK 304  
 DB 1005 SGNKPKSPVIVSKSGELIS---TSAKFSAGADSLYVSPBG---EESGEVYCTATNT 1056  
 QY 305 TGLKHQEVNLVY-----MRATQLOK-----NLTCFY-----W--- 331  
 DB 1057 AGVAKRKYQLTVVVRPVFGDLRGLSQDKPVEISVLAGEEVLTPCEVKSPLPPITTAKE 1116  
 QY 332 ----GPTSPK--LMLSLLENKEAKVSKREKPVVNLPEAGMOCCLSD--SGQV--LLES 382  
 DB 1117 TQLISPFSPRHTLPSPGSMKTEIRTS-----DSGMVLCVATINAGNVAQVXL 1165  
 QY 383 NIKVLPWTST-PVHPRASA-----LPAPPTGSALP 411  
 DB 1166 NVHVPKIQRGPKLKVQVQGRVDIPGNAQSTPLP 1200

RESULT 43  
 Q8K0X1 PRELIMINARY; PRT; 739 AA.  
 AC Q8K0X1;  
 DT 01-OCT-2002 (Tremblrel. 22, Created)

DT 01-OCT-2002 (Tremblrel. 22, last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, last annotation update)  
 DT Vascular cell adhesion molecule 1.  
 GN VCAM1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Strausberg R.;  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC029823; AAH29823.1; -.  
 DR MGD; MGI:98826; Vcam1.  
 DR GO: GO:0016020; C:membrane; IEA.  
 DR GO: GO:0016337; P:cell-cell adhesion; IEA.  
 DR InterPro: IPR003987; ICAM\_VCAM-1.  
 DR InterPro: IPR003599; Ig.  
 DR InterPro: IPR007110; Ig-Like.  
 DR InterPro: IPR003598; Ig\_c2.  
 DR InterPro: IPR003989; VCAM-1.  
 DR Pfam: PF00047; Ig; 5.  
 DR PRINTS: PR01472; ICAMVCAM1.  
 DR SMART: SM00409; IG; 5.  
 DR SMART: SM00408; IGC2; 5.  
 DR PROSITE: PS50835; IG\_LIKE; 5.  
 DR Immunoglobulin domain.  
 KM SEQUENCE 739 AA; 81329 MW; 8B66F3AAFD9D720E CRC64;  
 SQ

Query Match 6.1%; Score 142; DB 11; Length 739;  
 Best Local Similarity 20.2%; Pred. No. 0.031; Indels 100; Gaps 15;  
 Matches 79; Conservative 65; Mismatches 147;

QY 32 KKGDVELCTASQKSIQFHW--KNSNQI-KILNGSGFLTKGPKLNDRADSRSLMD 88  
 DB 237 QEGGATWMTCSSEGLPAPLIFMGRKLDNRYQL----- 270  
 QY 89 QGNFPLIINKLIEDSDTYICE---VEDQEEVOLLV-----FGLTANSDTLLHQGS 138  
 DB 271 SGNAIITLLAMRBEDSGVYVCEGVNLIGRDKALEVELVQEKPIVIDISPGSVAAQVGS 330  
 QY 139 LITLESPPSSPSVQCRSPRGKNIQ-----GKTLVSQLELDSGTWTCVQLQNK 191  
 DB 331 VVLTCAALGCDSPSFSWRTQDSPINGVVRNKGAKSTLVSSGVFEDSHYLCATCLOK 390  
 QY 192 KVEFKIDIVLAFQKASSIYVKKGEQVEFSPLA-----FTVEKL 232  
 DB 391 TLEKRTQVEVYSF-----PEDRVIKSGPLVHGRPVTVNCTVPNVYPRDHLIELL 441  
 QY 233 TSGGELMWAERASSSKSWI-TFDLKNKEVSVKRVTDPKLQ-MGKKL---PLHLTLPQ 286  
 DB 442 KG-----ETLMKKYFLEBWGISLTKIETTFIPIEDTGKSLVCLARHSGEME 493  
 QY 287 ALFQVAGSGNLTALAEAKGKLGQEVNLVMRATQLOK---NLTCFWSGTPSKMLSL 342  
 DB 494 SEBKQROSQVPLVNVAP-----KETTIVSPSPILIEBGSVNLTCSSDGIAPAKILWSR 548  
 QY 343 KLENKEAKVSKREKPVVNLN---PEAGMOC 370  
 DB 549 QLNGELQPLSENTTILTFMSTKRDDSGIYVC 579

RESULT 44  
 Q8HYV1  
 AC Q8HYV1; PRELIMINARY; PRT; 885 AA.  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, last annotation update)  
 DE Glycocylophosphatidylinositol anchor 1 protein (Fragment).  
 GN MDGAL.





```

GN I-CON.
OS Procamburus clarkii (Red swamp crayfish).
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC Astacidea; Cambaridae; Procamburus.
OX NCBI_TaxID=6728;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21423462; PubMed=1152946;
RA Fukuzawa A., Shimamura J., Takemori S., Kanzawa N., Yamauchi M.,
RA Sun P., Maruyama K., Kimura S.;
RT "Invertebrate connectin spans as much as 3.5 micrometer in the giant
RT sarcomere of crayfish claw muscle.";
RL EMBL J. 20:4826-4835(2001).
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL: AB055861; BAB64297.1; --
DR GO: GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro: IPR000577; FGGY_kin.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR008957; FN_III-like.
DR InterPro: IPR007110; IG_II-like.
DR InterPro: IPR003598; IG_C2.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00041; Fn3; 5.
DR Pfam: PF00047; Ig; 49.
DR Pfam: PF00018; SH3; 1.
DR SMART: SM00408; IGC2; 13.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00933; FGGY_KINASES_1; 3.
DR PROSITE: PS50835; IG_LIKE; 49.
DR PROSITE: PS50002; SH3; 1.
KM Immunoglobulin domain; SH3 domain.
SQ SEQUENCE 17352 AA; 1962348 MW; 4BA157BEC042E42D CRC64;

Query Match 6.0%; Score 140.5; DB 5; Length 17352;
Best Local Similarity 21.1%; Pred. No. 3.3;
Matches 92; Conservative 69; Mismatches 156; Indels 119; Gaps 20;

37 VELTCTASQKSIQIOWKNSQIKILNQSGFLTKGPSKLNDRADRSRLMDQGNPLII 96
16247 VERT---NEKATYTMH-KDEBEITENHERKLVSGKTR-----KLVY 16285
97 KNLKIEDSDTYICEVEDQKEVQLVFGLTAN-----SDTHLLQGSGLTLTLESPGSS- 150
16286 MEATLSDGEHYTCVLDGQECTAELTVRELPAEIVRKMKQDVSGKGRATWEVELTKGDV 16345
151 -----BSYQCR-----SPRKNIGQKGLTSLVQLELDSSGTWTCTVLDNQKVEFKID 198
16346 ITWYKDEVELRFSHDYQLSIDGK---VORLWYNQCQFEDSGTYRAVVGKSECSATLVKE 16401
199 IIVLAFOKASSIYKKEGE-----QVEFSFPLAFVTEKLTGSGELMW---QAEKAS 246
16402 LQV-----EGDSKSLPRQMDVNFKTDATFVE-ITKYEVKMLREGABLS 16447
247 SSKSWITFDLKNKEVSKRVLTQDPKLQ-----MGKKLPALHTLPQALPQVYAGSGLTL 299
16448 SEKVIIVKEKAKRILIVKVSQTDACEYSCVLGNLKTSCVLHVVRMETAP----- 16497
300 ALAKKIGKGLHENVLVKRAVTOQKNTCEV---WGPFSPLM-----LSLKLKNEK---A 349
16498 ---KLPKEHQKVEIVTKG---KQAVLVVPTATPTTPVYHKGQLLNTNTEKLP 16549
350 KVKREKRPV---WLVNPEAGMMQC-LLSDGQVLLSNIVLPTWSTPVYPRAS----- 399
16550 TISEQESITTKOVENIDCEHYRLKLCNDGAAVADPTLKILKPSQPGTPEMEVTNVS 16609
400 -----ALPAPPTGSAL 410
16610 VTLHMLSLPKEDGGRAL 16625

```

```

ID Q9VP08 PRELIMINARY; PRT; 403 AA.
AC Q9VP08;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE CG7166 protein.
GN CG7166.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Broctier P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downs M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster K., Gabrielian A.E., Garg N.S., Galbraith W.M., Glaeser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,
RA Kimball B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X.P., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milphins N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reihert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).

RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Man K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorssett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frick E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McInosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Paclet J., Paragag V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svitek R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.

```



RA Miera S, Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochik S.E., Smith C.D.,  
RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Celniker S.E.,  
RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,  
RA Ashburner M., Gehart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,  
RT "Annotation of Drosophila melanogaster genome";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN (4)  
RP SEQUENCE FROM N.A.  
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE FROM N.A.  
RA FlyBase;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE003595; AAF51754.2; -  
DR FlyBase; FBgn0037107; CG7166;  
DR InterPro; IPR001064; Crystal1n.  
DR InterPro; IPR008957; FN\_III-like.  
DR InterPro; IPR003599; IG\_1-like.  
DR InterPro; IPR007110; IG\_1-like.  
DR InterPro; IPR003598; IG\_c2.  
DR Pfam; PF00047; IG\_3.  
DR SMART; SMO0409; IG\_3.  
DR SMART; SMO0408; IGC2\_3.  
DR PROSITE; PS00225; CRYSTALLIN BETAGAMMA; 1.  
DR PROSITE; PS50835; IG\_LIKE; 3.  
KW Immunoglobulin domain.  
QO SEQUENCE 403 AA; 45747 MW; 429C1039F539F76D CRC64;

Query match 6.0%; Score 139.5; DB 5; Length 403;

Best Local Similarity 24.4%; Pred. No. 0.021;  
Matches 57; Conservative 39; Mismatches 97; Indels 41; Gaps 11;

Qy	27	VVLVGKGDVLELCTTASQKKSIOFHMKNSQIKLGNQSGFLTQSPKLNDRADRSRL	86
Db	33	KVIV---GETLELPKVNOLGSFVILWRKGS-----SVLTAGHKUT--RQRPKI	78
Qy	87	WDOGNFPLIKMLKIEDSDTYICEVEDQEE-----VOLLV-----FGLTASDTHLQOQ	137
Db	79	--VEDYNQINGVKTQDAGDYITCQGDQDENDDQYHVLVPLTALPHNGQYARKG	136
Qy	138	SLVLTLESPPGSSPV-----QCRSPGRKNIQGGKTLVSLSOLELQDSGTWTCVVLQNOK	191
Db	137	TVTLECKASGNPVPPTLIFWFKDDVFSGP--THLSDSTILENVDRNRHAGTVQCSADNGVK	194
Qy	192	-KVEFKDIVVLA---FOKASSIYKKEGEQVEPSFLATFVYKLTGSGELMWQ	241
Db	195	DRVSDMDIQLTTLSPREITVERKSWHASGAYVE---LVCTIAGDVNSEMLYQ	244

RESULT	50		
007763			
ID	007763	PRELIMINARY;	PRT; 398 AA.
AC	007763;		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DT	01-JUN-2003	(TREMBLrel. 24, Last annotation update)	
DE	2B4 protein precursor.		
DE	NMRK OR 2B4.		
GN	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_Taxid=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6;		
RX	MEDLINE=94044757; PubMed=8228228;		
RA	Mathew P.A., Garni-Wagner B.A., Land K., Takashima A., Stoneman E.,		
RA	Bennett M., Kumar V.;		
RT	"Cloning and characterization of the 2B4 gene encoding a molecule		

RT associated with non-MHC-restricted killing mediated by activated natural killer cells and T cells.";

RL J. Immunol. 151:5328-5337 (1993).

DR EMBL; L1903//; AAA16333.1; -.  
DR BIR: TA9443: TA9443

DR MGD; MGI:109294; Nmrk.  
DR InterPro; IPR003599; Ig.  
DR InterPro; IPR007110; Ig-like.  
DR SMART; SM00409; IG; 1.

FT SIGNAL	1	18	POTENTIAL.
SEQ	SEQUENCE	398 AA; 45132 MM; 566CDDACAE33525C	CNC64;
Query Match,	5.9%;	Score 138;	DB 11; Length 398;
Best Local Similarity	21.0%;	Pred. No. 0.026;	
Matches	83;	Conservative	64; Mismatches 147; Indels 102; Gaps 15

```

0Y 18 LLLPATOQ-----NKVLGKKGGTVELTCLASOKKSIQPFMKNSNQ-----IKILNQG 66
Db 12 LLLRAHQGDCEPDBSEBEVAVGSKPQVQIARPNIQTKQVAVGMKKTGESHKRIEL----- 67
0Y 67 SFLTKPSKLNDRADSRRLMDQGNFLLIHNKLIKEDSYIICEVEDQKEEY-----QLL 121
Db 68 NMYNDGSPSWNSVSPSDIYG-EDYGDPAFALISAKXLQDSGHYLLIETITNGAKVCNKNPQLL 126
0Y 122 VFGLTANSPTLLLOQGSITLTLBEPFGGSPBVO-----CRSRGKNI----- 163
Db 127 IL-----DHV-----ETPNLKAQMKPWTNGTCQFLSLCTVTKDNVSAFYWRGSLTISN 176
0Y 164 QGKGTLSVSOLELDQDSGWTTC-----TYLONOKKVEFKIDIVILA 203
Db 177 QNSTHWNENQIDASSLHTTICNVNSNRASMANHTLFTNGCCSVSPSNFFLDPGVITIVILV 236
0Y 204 FOXASSI-----VYKKEGQVEFFSPPLAFTVEKLTGSGELMWQAE---RASSSKSWITFDL 256
Db 237 TLPFGAITICFCVWTKKRQRQLOFSPKREPLTIIYEVKDSASANDQGCSCASSGSPSAVQBGD 296
0Y 257 KNKEVSVKRVTOQDPLQWKGKLPJLHLTLPOALPOYAGSGNLTALBAETGKJLHDEVNLV 316
Db 297 RGOEELDRRVSE-----VLEBOLPOOTFPGD-----RGTMYSMIQCKP 333
0Y 317 MRATQLOKUNLCEVWGPSPKMLSLKLENKEAVYS 352
Db 334 SDSTSQEK---CTVYSAVQSPRKSGSKKRKNQSYLS 366

```

```

RESULT 51
Q9NR99          PRELIMINARY;      PRT;  2828 AA.
ID   Q9NR99
AC   Q9NR99;
DT   01-OCT-2000 (TREMBLrel. 15, Created)
DT   01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE   Adlican.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RA   Crowl R.M., Luk D.;
RT   "Identification of the gene encoding Adlican, a novel protein
RT   expressed in human arthritic tissues.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR   EMBL; AF245505; AAF6402.1; -.
DR   HSSP; P56276; ITLK.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003598; Ig c2.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.
DR   InterPro; IPR003591; LRR typ.

```

DR Pfam; PF00047; Ig; 12.  
 DR Pfam; PF00560; LRR; 6.  
 DR Pfam; PF01463; LRRCT; 1.  
 DR SMART; SM00408; IGC2; 10.  
 DR SMART; SM00082; LRRCT; 1.  
 DR SMART; SM00013; LRRNT; 1.  
 DR SMART; SM00369; LRR\_TYF; 3.  
 DR PROSITE; PS50835; IG\_LIKE; 12.  
 DR Immunoglobulin domain.  
 SO SEQUENCE 2828 AA; 312291 MW; A18377D8554F1FE1 CRC64;

Query Match 5.9%; Score 138; DB 4; Length 2828;  
 Best Local Similarity 19.4%; Pred. No. 0.4;  
 Matches 103; Conservative 63; Mismatches 180; Indels 184; Gaps 21;

QY 10 LLLVLLALLPATQGNKV--VLGKKGDPVELTCTSSQKKSIOFHWKNSNQIKILGNQGS 67  
 DB LARLHVALLPVYIHOEKLENISLPGLSIHICHTKAAPLPSVRW-----VLG---- 2084  
 QY 68 FLTGPSKLNDRADSRSLMDQGN--FP--LIINKLKIEDSDTYICE---VEDQKEE 117  
 DB 2085 -----DGTQIRPSQFLHGNLFFVFNGLTIYINLAPKDSGRGECVAANLVGSARKT 2134  
 QY 118 VQLLVFGLTANS-----DTHLQGQSITLTLESPPGSSPSVQCRSPRGKNIQ----- 164  
 DB 2135 VQLVNQRAAANARITGTSPPRTDVRVYGGTLKIDCSASGCPWPRILWRLPSKRMIDALFSF 2194  
 QY 165 -----GKTLVSQLELDQSGTWTCYVLQNG--KYVEFKIDIVVLAFOKASSTVYKK 214  
 DB 2195 DSRKIVFANGTLVAVKSTVDKADGYLC-VARNKVGDYVVLKVDVW---KPAKLEHKE 2249  
 QY 215 EGE-----QVEFSPFLAFVTEKLTGSGELMWQERASSSKSWRT 253  
 DB 2250 ENHKKVFGYGGDLKVDCAVATGLPNEPILSMSLPDGSLVNSFWQSD-----DSGRTKRYV 2303  
 QY 254 FD---LKNKEVSVKRVTDQPKLOMGKKLPLHLTLPOALPOYAGSGNLTALBAKTKGLHQ 310  
 DB 2304 FNNGTLYFNEVGKRE-----EGDYTGFAENQVQKDEM 2335  
 QY 311 EVNLVVMRATQLOKNTL-----CEWVGPTSPKMLSLKLENKEAKYSKRE 355  
 DB 2336 RAVAKVAVTAPATIRKNTIYLAIVQVPYGDVTVTACEAGKEPMKPTW--LSPNKKIPIPTSSER 2394  
 QY 356 KPYW-----VLNPEAGMOCILSDSG-----QVLESNIKVLPTWSTVPVHPASAL 401  
 DB 2395 YQIVQDGTLLIOKAQKSDSGNTYCLVRNSAGDEKRTVWIVHWQ----- 2438  
 QY 402 PAPPTGSALPDPTQ-----ASALPDPPASALPVALAV 434  
 DB 2439 --PPKINGNPNPTTVREINAGSRKLLIDCKAEGIPTPRVLMFPEGVVL 2486

## RESULT 52

Q90Y38 PRELIMINARY; PRT; 1259 AA.  
 AC O90Y38;  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE Neural cell adhesion molecule 1L.  
 GN 1L1CAM.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Plazer M., Brenner V., Reichwald K., Wiehe T., Okeche A.,  
 RT "Comparative sequence analysis of the mouse L1cam locus and the  
 RT corresponding region of human Xg28.";  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF133093; AAF22153.1; -

DR HSP; P20241; ICFB.  
 DR MGd; MG1:96721; L1cam.  
 DR GO; GO:0007411; P:axon guidance; IMP.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR007110; IG\_1-like.  
 DR InterPro; IPR003598; IG\_c2.  
 DR Pfam; PF00041; fn3; 4.  
 DR Pfam; PF00047; Ig; 6.  
 DR SMART; SM00060; FN3; 3.  
 DR SMART; SM00408; IGC2; 5.  
 DR PROSITE; PS50835; IG\_LIKE; 6.  
 DR Immunoglobulin domain.  
 SO SEQUENCE 1259 AA; 140916 MW; 25743C039892A22F CRC64;

Query Match 5.9%; Score 137.5; DB 11; Length 1259;  
 Best Local Similarity 20.0%; Pred. No. 0.14;  
 Matches 73; Conservative 61; Mismatches 136; Indels 95; Gaps 16;

QY 34 GPTVLLTCTASQKKSIOFHWKNSNQIKILGNQGSFLTGPSKLNDRADSRSLMDQGNP 93  
 DB 346 GPTALDQVQGRPQPELTWR-----INGMSMETVNNKQKRI-EQGS-- 387  
 QY 94 LIINKLKIEDSDTYICEVEDQKEE-----VQLLVFGLTANSDTHL-LQGQSITL 141  
 DB 388 LILSN--VQSPDTWVTCSEARNQGLLANAYIVVQLPARILTDQNTQYMAVEGSTAYL 445  
 QY 142 TLESPPGSSPSVQCRSPRGKNI-----QGKTLVSQLELDQSGTWTCYVLQNGKV 193  
 DB 446 LCKAGAPVPSVQWLDEEGTTLQDERPPRYANGTSLRDLQANDTGRFCQANDQNNV 505  
 QY 194 EFKIDIVVLAFOKASSI-----VYKKEGEQVEFSPFLAFVTEKLTGSGELMWQERASS 247  
 DB 506 TILANLQV---KEAQIQGPRSALEKKGARTFCQASFDPSL---QASIMWRDGR-- 557  
 QY 248 SKSWITFDLKNKEVSVKRVTDQPKLOMGKKLPLHLTLPOALPOYAGSGNLTALBAKTKG 307  
 DB 558 -----DLOERGSDDKYFIEDGKLV-----QSL-DYSDQGYSCVASTERDE 598  
 QY 308 LHQEVNLVVMR-----TQLOKNTLCEVWGP---TSPKMLSKLENKE 348  
 DB 599 VESRQQLLVSPGVPVPHLESDRHLKQSOVHLS--WSPADHNSPIEKYDIEFEDKE 655  
 QY 349 AKVSK 353  
 DB 656 MAPKK 660

## RESULT 53

044924 PRELIMINARY; PRT; 1395 AA.  
 AC 044924;  
 DT 01-JUN-1998 (TReMBLrel. 06, Created)  
 DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)  
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)  
 DE ROUNDABOUT 1.  
 GN ROBO OR ROBO1 OR CG13521.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidae; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98117249; PubMed=9458045;  
 RA Kidd T., Brose K., Mitchell K.J., Fetter R.D., Tessier-Lavigne M.,  
 RT Goodman C.S., Tear G.;  
 RT "Roundabout controls axon crossing of the CNS midline and defines a  
 RT novel subfamily of evolutionarily conserved guidance receptors.";  
 RL Cell 92:205-215 (1998).  
 DR EMBL; AF040989; AAC38849.1; -  
 DR HSP; P56276; ITIK.  
 DR Flybase; FBgn0005631; robo.

DR GO:0005886; C:Plasma membrane; IDA.  
DR GO:0007411; P:axon guidance; IMP.  
DR InterPro: IPR003962; FNIII subd.  
DR InterPro: IPR003961; FNIII.  
DR InterPro: IPR008957; FN-III-like.  
DR InterPro: IPR007110; IG-III-like.  
DR InterPro: IPR003598; IG\_c2.  
DR Pfam: PF00041; Fn3; 3.  
DR Pfam: PF00047; Ig; 5.  
DR PRINTS: PR00014; FNIIIPEIIT.  
DR SMART: SM00060; FN3; 3.  
DR SMART: SM00408; IGC2; 5.  
DR PROSITE: PS00835; IG\_LIKE; 5.  
DR Immunoglobulin domain; Repeat.  
SQ SEQUENCE 1395 AA; 151778 MW; B020E234A5218983 CRC64;  
  
Query Match 5.9%; Score 136.5; DB 5; Length 1395;  
Best Local Similarity 21.1%; Pred. No. 0.2;  
Matches 102; Conservative 53; Mismatches 179; Indels 149; Gaps 18;  
  
QY 26 NKVVLGKKDPTVELTCTASQKSIQFHWKXSNQIKLGNQSFLLTKPSTLNDRADSR 85  
DB 353 NKRV-GING-VQDLPMAAGNPPPSVFWTK-----EGVSTLMFPNSSHGR 395  
QY 86 LMDQGNFPLIIKLIKEDSDTYICEVEDQKEVQLVFGLTANSDFH---LLQGSLTLT 142  
DB 396 QYVADDTLQITDRODEGEYVCASVSDSVTRVFLQVSSVDERPPIIIQIPANQT 455  
QY 143 LESPPGSSPSVQCR-----SPRGK-----NIQGGKTLVSQLELDSDGTWT 183  
DB 456 L--PKGSVATLPCRATGNPSPRIKMFHGHAVQAGNRYSIIGSSSLRVDLQLSDGTYT 513  
QY 184 CTYLNQKKEFKIDIVYLAFOKASSIVYKKEGQVFFSPLAFYTKLTGSELMWQAE 243  
DB 514 CTA-----SSRGSTSWAATLYVEK--PGSTSLRRPAD 544  
QY 244 -----RASSKSWITFDLKNKEVS--VKRVTD---PKLQMGKPLH- 281  
DB 545 PSTYPAAPGTPKVLNVSRITSLRMKASQKRPAGVGIIGTYVEYFSPDQITQVIYAAHR 604  
QY 282 -----LTLPAALPOY-----AGSGNLTLALEA 303  
DB 605 VGDQVITISGLTPTGSYVFLVRAENTQGISVPSGLSNVITKIEADPDASANDLSAARTL 664  
QY 304 KTKGLHDEVLVVNRATQLOKNTLCEYWGPTSPKMLSLKLNKEAIVSREKRVWYVLP 353  
DB 665 LTKSVELLIDASAINASAVLEWMLHV--SADEKYVEGLRIHYKDSV-----P 711  
QY 364 EAGMWCCLLSD--SGQVLESINIKVLPWSTFVHPRASALPAPPTGSLPDPOTASALPDP 422  
DB 712 SAQYHSTTVWDASAESEFVGNLKKYTKYEFPLTPPFETIEGQPSNS-----KTLATYEDV 766  
QY 423 PAA 425  
DB 767 PSA 769  
  
RESULT 54  
Q770G5 PRELIMINARY; PRT; 1465 AA.  
AC Q770G5;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Scurionath; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haile F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.J., Scheetz T.E.,  
RA Brownstein M.J., Uedlin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Greenwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywiński M.I., Skalska U., Smalusz D.E., Schmech A., Schein J.E.,  
RA Jones S.J., Maier M.A.,  
RA "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Brain;  
RA Strausberg R.;  
RA Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC054540; AAH54540.1; --  
KW Hypothetical protein.  
SQ SEQUENCE 1465 AA; 159967 MW; AB51C378BC3BA32 CRC64;  
  
Query Match 5.9%; Score 136.5; DB 11; Length 1465;  
Best Local Similarity 22.8%; Pred. No. 0.21;  
Matches 115; Conservative 53; Mismatches 191; Indels 145; Gaps 27;  
  
QY 10 LILVLOLALL-----PAATQKNV-----VLGKKGVVELT 40  
DB 25 LILLLPLLLLLGRPASGAATKSGSPQAGASVTRTFPYFLVEVDTLVSGSSVILN 84  
QY 41 CTASQKSIQFHWKXSNQIKLGNQSFLLTKPSTLNDRADSRSLMDQGNFPLIIKNL- 99  
DB 85 CSAVSESPSPNIEK-----KDGTF-----LNLSDDRRLPPOS--LFLSNVY 126  
QY 100 ----KIEDSDTYIC--EVED---QKEVQLVFGI---TANSDFHLLQGSLLTLTDS 145  
DB 127 HSKMKNRPDGBFQCVATVNLGITVSRTAKLTVAGLPRTSQPEPSVYVGNALINCEV 186  
QY 146 PGGSSPSV---QCRSP-----RGKNIQGGKTLVSQLELDSDGTWTCTYLN-----QKK 192  
DB 187 NADLVFVFWEQNRQPLLDRIYKLPSC--TLVISNATGEGDGLYRCIVESGGPPKFSDE 245  
QY 193 VEFKI-----DIVYLAFOKASSIVYKKEGQVFFSPLAFYTKLTGSELMWQARAS 247  
DB 246 AELKVLQDPBEIVDLVFLMRPSSMKRVTGQ---SAVLPCVVSGLPAPVVRW----- 293  
QY 248 SKSWITFDLKNKEVSVKRVTDPPKLOMGKPLPHLTLTPALPOY---AGSGNLTL----- 298  
DB 294 -----MKNEEVLDTSSGRLVLLAGGLEISDVTEDDAGTYFCIANDNGKNTVAQAE 345  
QY 299 LALEATGKLHDEVLVNRATQLOKNTLCEYWGPTSPKMLSLKLNKEAIVSREKRV 354  
DB 346 LTVQVPPGFLKQAPANIYAHESMDIV--FCEVATGKPTPVKWKNDGVVIPSDFIXIVE 403  
QY 355 E--KPVWLVNPEAGMWCCL--SDSGO-----VLESINIKVLPWSTFVHPRASALPAP 404  
DB 404 HNLQVGLVKSDEGFQCTAENDVGAQAQAQILTIHDV-AIPT-----LP 449  
QY 405 PTG--SALPD---POTASALPDP 423  
DB 450 PLSUTSATTDHLAPATGFLPSAP 473  
  
RESULT 55  
Q08IRV9 PRELIMINARY; PRT; 4117 AA.  
ID Q08IRV9





```

SQ SEQUENCE 4179 AA; 461782 MW; 47804277D0914E63 CRC64;
Query Match 5.9%; Score 136.5; DB 5; Length 4179;
Best Local Similarity 17.8%; Pred. No. 0.91;
Matches 61; Conservative 61; Mismatches 109; Indels 111; Gaps 13;

QY 19 LPAATGKNKVLG-----KKGDTVELTCTASQKSIQPHWKN-----SNOIKILGNQ 66
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3291 LPAKSDYSLKLDQSSNLKRGESTVECEYSSDDTYTDVWHESSDCAPISSNNVROQGN- 3349
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 67 SFLTKPSPKLNDRADRSRLMDQGNPEPLIKLKIDSDTYICEVDQKEBVLVGLT 126
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3350 -----LVISNPSPDAGNYCKCKTIDGDLTYTTSYKLE 3382
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 127 ANSDTLHQGSLTLTLESPGSSPSVOC-----RSRPGCK-----NIOGKT-----LS 170
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3383 VEQPPELKSCKI-----VYAKVGNADIQCADESRQPTTRWSKQYQOLQAGSLNNEKIS 3439
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 171 VSQLELDQSGTWTCTV-LQNKQKVEFKIDIV-----LAFQKASSIVYKKE 215
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3440 LDSVQANDAGTYICTAQVADGETADFPNLLVTGAIQPRQEDRSYMSPTLPNSGFK-- 3497
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 216 GEVEESFPLAFVEKLTSGSGELWQAEKASSSKSWITTDKNN----- 259
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3498 -----ENFELTFRPE--NGDGLLFNGQTRSG-DYIALSLKDRYAEFRFDEGKEMLVR 3549
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 260 -----EVSVKRVTDPLQMGKLPPLHLTPOLAPQ 290
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 3550 AEEPLALNEMHYVRVSRFRKDGTYIQVDEGHPVAFPTLQIQIPQ 3591
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

RESULT 57
Q8MPN3 PRELIMINARY; PRT; 4223 AA.
AC Q8MPN3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Perlcan (CG7981-PD).
GN TROL OR EG:BACR2583.11 OR CG7981.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Voigt A.;
RT "perlcan participates in proliferation activation of quiescent
RL Drosophila neuroblasts.";
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Gabor G.L.,
RA Abiri J.F., Agayanti A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballou R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolhakov S.,
RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Evansgela C.C., Ferraz C., Ferreira S., Fischmann W.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasner K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,

```

```

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasco P., Lei Y., Lewitcky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mervet G., Mishina N.V., Modarri C., Morris J., Mohtrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spirding A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodger, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zhang L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorssett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Mohtrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacle J.B., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svitskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochick S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Milburn G., Richter J., Russo S.,
RA Seale S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Flybase;
RX SUBMITTED (SEP-2002) to the EMBL/GenBank/DBJ databases.
RA EMBL; AJ487018; CAD31650.1; -
RA EMBL; AE003424; AAN09079.1; -
RA Flybase; FBgn001402; trol.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0005198; P:structural molecule activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR001525; C5 DNA meth.
DR InterPro; IPR008985; ConA like_1ec_g1.
DR InterPro; IPR00742; EGF 2.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR003599; I9.
DR InterPro; IPR007110; I9-1like.
DR InterPro; IPR003598; I9_c2.
DR InterPro; IPR000034; Laminin B.
DR InterPro; IPR002049; Laminin B.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR000504; RNA_rec_mot.

```

```

DR Pfam; PF00047; Ig; 13.
DR Pfam; PF00052; laminin_B; 3.
DR Pfam; PF00053; laminin_EGF; 2.
DR Pfam; PF00054; laminin_G; 3.
DR Pfam; PF00057; 1d1_recept_a; 23.
DR PRINTS; PR00261; LDLRECEPTOR.
DR Prodom; PD003031; Laminin_B; 3.
DR SMART; SM00180; EGF_lam; 8.
DR SMART; SM00409; IG_13.
DR SMART; SM00408; IGF2; 13.
DR SMART; SM00281; Lamb; 3.
DR SMART; SM00282; Lamg; 23.
DR SMART; SM00192; LDLR; 3.
DR PROSITE; PS00094; C5_MCASE_1; 1.
DR PROSITE; PS00023; EGF_1; 10.
DR PROSITE; PS01166; EGF_2; 6.
DR PROSITE; PS01835; IG_LIKE; 12.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
DR PROSITE; PS00205; LAM_G_DOMAIN; 3.
DR PROSITE; PS01209; LDLR_1; 20.
DR PROSITE; PS00068; LDLR_2; 23.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Immunoglobulin domain.
SQ SEQUENCE 4223 AA; 466919 MW; 31D7C2B5C9B1D2E0 CRC64;

Query Match 5.9%; Score 136.5; DB 5; Length 4223;
Best Local Similarity 17.8%; Pred. No. 0.92; Matches 61; Mismatches 109; Indels 111; Gaps 13;
Matches 61; Conservative 61;

QY 19 LPAATQGNKVLG-----KKGDVELTCTASQKKSIOFMKN-----SNQIKILGNQG 66
DB 3335 LPAKSRYSYKLDQSSNLAGESTDYECVSSDDTYTDVWENSDBGAPLSNNRYOVGN- 3393
QY 67 SFLTKGSKLNDRAADRSRLMDQGNFPLIITKLTIEDSDTYICEVEDQKEEVOLLVFGLT 126
DB 3394 -----LVISNVSPPDAGNVYCKCKTDEGDLTYTSYKLE 3426

QY 127 ANSDTHLLOGQSULTLESPPGSSPSVQC-----RSPRGK-----NIOGKKT-----LS 170
DB 3427 VEDDPHELKSSKI---VYAKVGAADQCCAGADESRQPTTYSWROYGOLQGRSLMEKLS 3483
QY 171 VSQLELDQSGTWCTV-LQNKQKVEFKIDIV-----LAFQKASSIVYKKE 215
DB 3484 LDSQANDACTICTAQAAYAGETADFNILVTGAIQFQRPQSPRSVMSFTLPSNSFK- 3541
QY 216 GEQVESEFPPLAFYVEKLTGSGELMWQERASSSSKWTTPDKNK----- 259
DB 3542 -----FNFELTFRPE--NGDGLLIFNQTRGSG-DYIALSLKDRYAEFRFDGKPMVLR 3593
QY 260 -----EVSVKRVTODPKLQMGKXLPKHLTLPQALPQ 290
DB 3594 AEEPLALNEMHTVAVSRFKRDGYIQVDEQHPVAFPTLQQLPQ 3635

RESULT 58
Q8IRV8 PRELIMINARY; PRT; 4228 AA.
AC Q8IRV8;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE CG7981-PE.
GN TPOL OR EG: BACR25B3.11 OR CG7981.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Cealnikher S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

```

```

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abiri J.F., Agdayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman J.P., Bhandari D., Bolhakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ileguam C.,
RA Jalali M., Kalush F., Kapran G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kul D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshirei A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murzy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodgett, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RA Science 287:2185-2195(2000).
[2]
RP SEQUENCE FROM N.A.
RA Cealnikher S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorett V., Doup L.E., Doyle C., Drenek D., Farfan D.,
RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegam C., Jalali M., Kruse D., Li P., Matrei B., Moshirei A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleab J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phoumenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RA "Sequencing of Drosophila melanogaster genome."
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby Y.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Cealnikher S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungell C.J., Lewis S.E.;
RA "Annotation of Drosophila melanogaster genome."
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Cealnikher S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE FROM N.A.
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RA EMBL; AE003424; AAN09078.1; -

```

```
DR FlyBase; FBgn0001402; trol.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR001525; C5_DNA_meth.
DR InterPro; IPR008985; Cna1_like_g1.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR000034; laminin_B.
DR InterPro; IPR002049; laminin_EGF.
DR InterPro; IPR001791; laminin_G.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR000504; RNA_rec_moc.
DR Pfam; PF00047; Ig; 12.
DR Pfam; PF00052; laminin_B; 3.
DR Pfam; PF00053; laminin_EGF; 2.
DR Pfam; PF00054; laminin_G; 3.
DR PRINTS; PR00261; LDLRECEPTOR.
DR ProDom; PD003031; laminin_B; 3.
DR SMART; SM00181; EGF; 9.
DR SMART; SM00409; EGF_lam; 6.
DR SMART; SM00408; IGC2; 12.
DR SMART; SM00281; LamB; 3.
DR SMART; SM00282; LamG; 3.
DR SMART; SM00192; LDLa; 24.
DR PROSITE; PS00094; C5_MTASE_1; 1.
DR PROSITE; PS00022; EGF_1; 10.
DR PROSITE; PS01186; EGF_2; 6.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
DR PROSITE; PS01209; LAM_G_DOMAIN; 3.
DR PROSITE; PS01209; LDLRA_1; 21.
DR PROSITE; PS00068; LDLRA_2; 24.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 4228 AA; 466934 MM; 0F630AA0BBF4CD33 CRC64;

Query Match 5.9%; Score 136.5; DB 5; Length 4228;
Best Local Similarity 17.8%; Pred. No. 0.93;
Matches 61; Conservative 61; Mismatches 109; Indels 111; Gaps 13;

QY 19 LPAATGAKVYLG-----KKGDTVELTCTASQKSIQFHWKN-----SNQIKILGNQ 66
DB 3340 LPAKSRDYSLKLDQSSNLPAGESTDVECYSDDTYTDVWERSDGAPLSNVNRQVGNR- 3398
QY 67 SFLTGPSKLNDRADSRSLMDQGNFPLIKNKIEDSDTYICEVEDQKEVQLVFGLT 126
DB 3399 -----LVLSNVSPSDAGNYVCKCKTDEDDLTTTSYKLE 3431
QY 127 ANSDTHLQGSLLTLLESPGSSPSVQC-----RSRPGK-----NIQGGT-----LS 170
DB 3432 VEDQPHLKSSKI---VYAKVGANADLQCGADESRQGYRWISQYGLQGRSLMMKIS 3488
QY 171 VSQLELDQSGTWCTV-LQNKQVEFKIDIVY-----LAFQKASSIYKKE 215
DB 3489 LDSVQANDACTYICTQYADGETADFPNLIIVTGAIPOFROBRSYMSFPTLPSNSSFK- 3546
QY 216 GEQVEFSPLAFVVEKLTGSGELMMQAEERASSSKSWITPDLKNK----- 259
DB 3547 -----INFELTFPE--NGDGLLFGNGQTRGSG-DYIALSLKRIYAEFRPDGKMLYR 3598
QY 260 -----EVSVKRVTDPRKIQMGKKLPLHLTLFQALPQ 290
DB 3599 AEEPLALNEMHTVRSFKRDGYIQVDEGHPVAFPPTLQQLPQ 3640

RESULT 59
Q96DN8
```

```
ID Q96DN8 PRELIMINARY; PRT; 512 AA.
AC Q96DN8;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein FLJ31774.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ieshbaeni T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA Houta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA Sugiyama T., Irie R., Oseuki T., Sato H., Wakamatsu A., Ishii S.,
RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K., Isogai T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK056336; BAB71154.1; -.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR Pfam; PF00047; Ig; 6.
DR Pfam; PF00048; IGC2; 5.
DR PROSITE; PS00835; IG_LIKE; 6.
KW Hypothetical protein; Immunoglobulin domain.
SQ SEQUENCE 512 AA; 54971 MM; 9368150E8A5BD6C8 CRC64;

Query Match 5.8%; Score 136; DB 4; Length 512;
Best Local Similarity 20.9%; Pred. No. 0.053;
Matches 86; Conservative 48; Mismatches 150; Indels 128; Gaps 17;

QY 34 GDTVELTCTASQKSIQFHWKNSNQIKILGNQSFVTGSPSKLNDRADSRSLMDQGNF 93
DB 148 GSNVTLPCVQGYPRPTIKMRLDNMPFSR--PFSVSSISQLRTGA----- 192
QY 94 LIKILKIEDSDTYICEVEDQ-----KEEVQLVPLGLT-----NSDTHLQGSLLTL 143
DB 193 LFTLNLMSDKGTIYCEANQFGKIQSETTVVTGLVPLIGISPSVANVILGQQLTLP 252
QY 144 ESPSPS-----SPSVQCRSPRGNKIQGGTSLVSQLELDQSGTWCT--- 184
DB 253 TLLAGNPPIPERMINSAMLLQNPYITRS-----DGSIIIRVQDQDGETCVAS 304
QY 185 TYLQNKQVEFKIDIVLAFQKASSIYKKEGEQVEFSFPLAFVVEKLTGSGELMMQAE 244
DB 305 NVAGTNNTKTSVVVHVLPTIQHQOILSTIEG--IPVTLF-----CKA 345
QY 245 ASSSKSWITPDLKNEVSVKRVTDPRKIQMGKKLPLHLTLFQALPQYAGSGNLTLLAEK 304
DB 346 GSNPRPSYVWSKKGLIS---TSAKPSAGADSLVYVSPG---EESGEYCTAINT 397
QY 305 TCKLHQEVLVYM---RATOLQKNLTCEVWGPTSPKMLSLTENKKAQVSRKRPVWL 361
DB 398 AGYAKRKQVLYTVAPRFRFGDQRLS-----QKXPVVIS-----VL 433
QY 362 NPEAGMOCCLSDSGOVLIESNIKVLP---TWS-----TPVHPRASALPA 403
DB 434 AGE-----EVLTPCEVKSLLPPIITWAKETOLISPSRPHFTPLS 473

RESULT 60
Q9VCT4
ID Q9VCT4 PRELIMINARY; PRT; 545 AA.
AC Q9VCT4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE KLG protein (LD10776p).
SQ KLG OR CG6669.
```



OC Drosophila melanogaster (Fruit Fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peerygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxId=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkelley;  
RC MEDLINE=20196006; PubMed=10731132;  
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amandides P.G., Scherer S.E., Li P.W., Hoekne R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mkios G.L.G.,  
RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
RA Ballew R.M., Baabu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolishkov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brodtier P.,  
RA Buttle K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport I.B., Davies P.,  
RA de Pallos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris W.,  
RA Hardie N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hootin D., Houston K.A., Howland T.J., Wei M.-H., Ibbagwan C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lesko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Schaefer F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svyrkas R., Tecor C., Turner E., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinlock G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhu Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gbbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*.";  
RT Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkelley;  
RC STRAIN=Berkelley;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Flise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nuno J., Pacle J., Parages V., Park S., Phouenavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AE003741; AAF56071.1; -  
DR EMBL; AY060363; AAL25402.1; -  
DR FlyBase; FBgn0017590; klg  
DR InterPro; IPR0003961; FN\_III  
DR InterPro; IPR008957; FN\_III-like.  
DR InterPro; IPR007110; IG\_1like.  
DR InterPro; IPR003598; IG\_c2.  
DR Pfam; PF00047; Ig; 3.  
DR SMART; SMO0060; FN3; 1.  
DR SMART; SMO0408; IGc2; 3.  
DR PROSITE; PSS0835; IG\_LIKE; 3.  
KW Immunoglobulin domain.  
SQ SEQUENCE 545 AA; 60087 MW; D5E81D9E5574E9DE CRC64;

Oy	2	NRGPPFPHLLLVLTAL-----LPAATGKNVVLGKSGDYTELTCTASQKKSIGFW	53
Dd	75	NRGNSNSMSNVQOSAVALASTLTTLTREFLSRGHTTAAVVGDTLVLPQVENLGNFVLLM	134
Oy	54	KNSNOIKILGQGSFLLTKGSPSKLNDRADSRRLSDGQNFPLIKMLKIEDSDTYICEV--	111
Dd	135	RRGNVLTASN--IMVTR-----DERVRLID--GYNLEISDLEPQADGVVCGQSD	181
Oy	112	---EDQKEEVOLL-----VFGLTNSDTHLLQGSLTTLSPGSSPSVQCSFRGN--	162
Dd	182	KINRDQVHTVIELVPPSVRALPTSGOLQARKGSGITLLECKSGNPPSVISYWTKKSGANKS	241
Oy	163	---TGGKTVLSVSLTSDSGTWTCV---LQNKVPEFKDIDVLA--FOKASSIVYKKE	215
Dd	242	TARIGDGFILTEKLEKQAGVCTCTDNGVGDVYVDMRLDVLVPPDIQVEXKWHISGE	301
Oy	216	GEQVEF-----SPLAFTVEKLTSGSELMWOAERASSSKSWITF--DLK	257
Dd	302	GFEAKLVCIYPADPVATVWYQNSFFLGSTDRRI-----MATRA--NRHMLTIRHIQ	351
Oy	258	NKEVSVKRVTDPLQLQMGKPLRLTLTPALPOVA-----GGNLTTLAEATGKLHQ	310
Dd	352	QEDGKNYSQVADNSLIGRSRKTMELSGRGAEFYSPKMGSRPDSYNTLWKIDSYPL--E	409
Oy	311	EVNIVNVRATQLOKNTLCEVWG-----PTSPKL--MELSKLEN-----KEAK	350
Dd	410	EVRLILYR--VQNNETYYQGRNRHDLILTPENRPAEBPLTHINSYTIKNLHPGGYEA	466
Oy	351	VSKEKEPVVILNPPAGMOCCLSDSGVLLNESIKVLPTWSTPVHPASALPAPTGSAL	410
Dd	467	VQAKNRXGM--NEVSDIFQFVVAITNSQDIGEDAELVA--SSKRSNSASIGSLPGSTVL	522
Oy	411	PDPTASALPDPASALPALANISF--LLGGG	442
Dd	523	-----HTALDLALALSNCPDRRLTGG	545
RESULT 61			
O9NEG1		PRELIMINARY; PRT; 1035 AA.	
AC	O9NEG1		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE	EG-BACR25B3.11 protein.		
GN	TROL OR EG-BACR25B3.11 OR CG7981.		
OC	Drosophila melanogaster (Pruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;		
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Murphy L., Harris D., Barrell B.;		
RT	"Sequencing the distal X chromosome of Drosophila melanogaster."		
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Benos P.;		
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AL138972; CAB72284.1; --.		
DR	HSSP; P00740; IEDM.		
DR	FLYBase; FBgn0001402; trol.		
DR	GO; GO:0003676; F:nucleic acid binding; IEA.		
DR	InterPro; IPRO08985; ConA_like_1ec_g1.		
DR	InterPro; IPRO00742; EGF_2.		
DR	InterPro; IPRO06209; EGF_like.		
DR	InterPro; IPRO06210; IEGF.		
DR	InterPro; IPRO07110; Ig-like.		
DR	InterPro; IPRO03598; Ig_c2.		
DR	InterPro; IPRO01791; Laminin_G.		
DR	InterPro; IPRO02172; LDL_receptor_A.		

DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00008; EGF; 3.  
 DR Pfam; PF00047; Ig; 2.  
 DR Pfam; PF00054; laminin\_G; 3.  
 DR Pfam; PF00057; Id1\_recept\_a; 1.  
 DR SMART; SM00181; EGF; 2.  
 DR SMART; SM00408; IgC2; 2.  
 DR SMART; SM00282; LamG; 3.  
 DR SMART; SM00192; LDLa; 1.  
 DR PROSITE; PS00022; EGF\_1; 3.  
 DR PROSITE; PS01186; EGF\_2; 2.  
 DR PROSITE; PS50835; IG\_LIKE; 2.  
 DR PROSITE; PS50025; LAM\_G\_DOMAIN; 3.  
 DR PROSITE; PS01209; LDLRA\_1; 1.  
 DR PROSITE; PS50068; LDLRA\_2; 1.  
 DR PROSITE; PS00030; RRM\_RNP\_1; 1.  
 DR EGF-like domain; Immunoglobulin domain.  
 SQ SEQUENCE 1035 AA; 114044 MW; 413DE7EDD3AE711 CRC64;

Query Match 5.8%; Score 134.5; DB 5; Length 1035;  
 Best Local Similarity 17.6%; Pred. No. 0.19;  
 Matches 57; Conservative 58; Mismatches 103; Indels 105; Gaps 12;

QY 32 KKQDTVELTCTASQKSIQPHMK-----SNOIKILGNQSGFLTKGPSKLANDRADSRSS 85  
 DB 179 RAGESTIDVECYSSDDTYTDVWERSDGAFLSNVNRQVGNR----- 218  
 QY 86 LMQGNPRLIKLKLIEDSDTYICEVEDQEEVQLVFGLTANSDTHLQGSLLTLLES 145  
 DB 219 -----LVISNVSPSDAENVCKCTDEGDYTTTSYKLEVEDQPHLESSKI---VYA 267  
 QY 146 PGSSPSVQC-----RSPRGK-----NIOGKT-----LSVQLELDSDGTWCTV-IQ 188  
 DB 268 KVGANNDLQCGADESRQPTWRMSROYQLQAGRSIMNKKSLSDSQANDAGTICTAQA 327  
 QY 189 NQKKVEFKIDIV-----LAFQKASSIVYKKEGQVBSFPFLATVEKLTG 234  
 DB 328 DGETADFPNLLVVTGALPQFROPSPRSYSPFLTPNSSEF-----FNELTFRPE--NG 378  
 QY 235 SGEIMQAERASSSSKITFDLKNK-----EVSRYRYT 267  
 DB 379 DGLLFNGQTRGSG-DYIALSLDQRYAEFFPDGKPMVLRAEPLALNEMHTTVSRFK 437  
 QY 268 QDPKLGWKKLPLHLTLPLQALPQ 290  
 DB 438 RDGYIQVDEQHPVAFPLQLQIPQ 460

RESULT 62

Q772H2 PRELIMINARY; PRT; 487 AA.  
 AC 0772H2;  
 DT 01-OCT-2003 (Tremblrel. 25, Created)  
 DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Fibroblast growth factor receptor-like protein precursor.  
 GN FGFR1.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OC NCB1\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Cartilage;  
 RA Tsub B., Zhuang L., Taeschler S., Wiedemann M.,  
 RT "Characterization of FGFR1, a Novel FGF receptor preferentially  
 expressed in skeletal tissues.",  
 RL J. Biol. Chem. 0:0-0(2003).  
 DR EMBL; AJ535114; CAD59380.1; -.  
 DR Receptor; Signal.  
 KM SIGNAL 1 18 Potential.  
 FT CHAIN 19 487 fibroblast growth factor receptor-like

FT SEQUENCE 487 AA; 54099 MW; FFD0132AD917FF94 CRC64;  
 SQ

Query Match 5.8%; Score 134; DB 13; Length 487;  
 Best Local Similarity 21.2%; Pred. No. 0.071;  
 Matches 89; Conservative 55; Mismatches 149; Indels 126; Gaps 19;

QY 117 EVQLVFGITLNSDT-----HLQGSV-----TLLESP-PGSSPSVQCRSPRGKN 162  
 DB 4 QALLLAGIVLSDSARGPRIADKVIHQSVRLRTIKLCFVGDDPPLTMKKDGT 63  
 QY 163 IQGKTI-----LSVQLELDSDGTCTVQLNQKVEFKIDIVLAFAQASSIVYKK 214  
 DB 64 IHSGWTFRLIQGGLKIKVESESDAGTYICATNGFGSTNNVYTLIVIDDTSCKNSQTP 123  
 QY 215 EEEQVEFSPLAFTYBKLTGSGELMWQ-----AERASS-----K 249  
 DB 124 EESNGEYE-----DHSQKMAQPPFTPOPAKRRRRIARPVGSSIRLKVASGNPR 173  
 QY 250 SWTFDLKKNKEVSVRVTQDPKLGWKKLPLHLTLPLQALPQYAGSGNLTLLAATGKLH 309  
 DB 174 PDLTKKKNKPLMPHEIGEN-----KKKWTNLNKNLKE--DSGKTCRVFNKVGELN 225  
 QY 310 QEVNLVVRATQLQK-----NLTCVWGPFSFKMLSLKLENKAVSKREKPV--W 359  
 DB 226 ATYKVEVLIQRTSKPDLTGTHPNTTVYGGTTS-----FOCKVRSDKRVIQW 274  
 QY 360 VLNPEAG--WQGLSDSGGVLTLESNIKVLPT-----WSPN-----VHPRAS----- 399  
 DB 275 LKRVVEGSKSYNSTIIVDGGQKIV-----VLPTGVSWSRSDGSYLNKMLITRAKEDAGM 329  
 QY 400 --ALPAPPTG-----SALPDPQTASALPDPAPASALPALAVISPLLAGLVACV 448  
 DB 330 YICLGNTMGYSFRSAPLTLVLPDPKPSAPVPSPSSVSLPMWV-----IIGIPAGAVFI 383

RESULT 63

Q63155 PRELIMINARY; PRT; 1445 AA.  
 AC 063155;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Colorectal tumor suppressor.  
 GN DCC.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCB1\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97015074; PubMed=8861902;  
 RA Keino-Masu K., Masu M., Hinck L., Leonardo E.D., Chan S.S.Y.,  
 RA Culicci J.G., Tessier-Lavigne M.,  
 RT "Deleted in Colorectal Cancer (DCC) encodes a netrin receptor";  
 RL Cell 87:175-185(1996).  
 RN [2]  
 RP SEQUENCE OF 387-420 FROM N.A.  
 RX MEDLINE=90100559; PubMed=2294591;  
 RA Fearon E.R., Cho K.R., Nigro J.M., Kern S.E., Simons J.W.,  
 RA Ruppert J.M., Hamilton S.R., Preisinger A.C., Thomas G., Kinzler K.W.,  
 RA Vogelstein B.,  
 RT "Identification of a chromosome 18q gene that is altered in colorectal  
 cancers";  
 RL Science 247:49-56(1990).  
 DR EMBL; U68725; AAB41099.1; -.  
 DR EMBL; M32291; AAA41086.1; -.  
 DR HSSP; P56276; ITLK.  
 DR InterPro; IPR003962; FNIII subd.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR007110; Ig-1-like.  
 DR InterPro; IPR003598; Ig\_c2.

DR Pfam: PF00041; fn3; 6.  
 DR Pfam: PF00047; Ig; 4.  
 DR PRINTS: PRO0014; ENTYPEIIT.  
 DR SMART: SM00060; FN3; 6.  
 DR SMART: SM00408; IGC2; 3.  
 DR PROSITE: PS50835; IG\_LIKE; 4.  
 KM Immunoglobulin domain; Repeat.  
 SO SEQUENCE 1445 AA; 157940 MW; 084F625954481988 CRC64;

Query Match 5.8%; Score 134; DB 11; Length 1445;  
 Best Local Similarity 20.7%; Pred. No. 0.32;  
 Matches 75; Conservative 48; Mismatches 132; Indels 108; Gaps 14;

DR 34 GDTVELICTASQKSIQFHKNKSNQIKILGNQSFILTKGSKIND-RADSRRLMDQGNF 92  
 DB 154 GDTVLLKCEVIGDPMPTIHWQKNQ-----DLPPIFGDSRVVVLPSG-- 195  
 QY 93 PLIIKNIKIEDSDTYICEVED-----QKEEVL-----VFGLTANSPTHLIQQ 137  
 DB 196 ALQISRLQPDSDGYRCSANNPASTRTGNFAEVRILSDPGLHQLVFLQRPNVIAIEGK 255  
 QY 138 SLTL---TLSPSPG-----SPSVQCRSPRKNIQCGKTLVSQLELDQSGTWCTVLQ 188  
 DB 256 DAVLECCVSGYPPSPFTMLRGEVYQLRSKK-VSLGSGNLLISNVTDDSGTYTCVVTY 314  
 QY 189 NQKVEKIDIVLA---FQKASIVYKKEGQVEFPFLAFYVEKLTGSGELMWQERA 245  
 DB 315 KNEITSAEELTVLPWPLNHPNLSLAYESMDIEF-----ECA 353  
 QY 246 SSSKSWITFD-LKNKEVSVKRVTDPKLQMGKLLPLTLPLQALPYAGSGNLT----- 299  
 DB 354 VSGKRPVTVVMKNGDY-----VPSDFQVGSNLRILGVK 392  
 QY 300 -----ALBAKTGKLHQEVLVVMRATQLQKNLTCEWGPPTSFKMLSLKLENKAVY 351  
 DB 393 SDEGFYQCVANENKGNASSAQLIVPKPAIPSSSIL-----PSAPRVVPLVSSRFVRL 447  
 QY 352 SKR 354  
 DB 448 SWR 450

RESULT 64  
 Q96AA2 PRELIMINARY; PRT; 6620 AA.  
 AC Q96AA2;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Obscurin.  
 OS Obscn.  
 GN Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Heart;  
 RX MEDLINE=21342081; PubMed=11448995;  
 RA Young P.W., Ehler E., Gaulel M.;  
 RT "Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor  
 RT protein involved in sarcomere assembly.";  
 RL J. Cell Biol. 154:123-136(2001).  
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
 DR EMBL: AJ002535; CAC44768.1; -.  
 DR Genew; HGNC:15719; OBSCN.  
 DR GO: GO:0005524; F:ATP binding; IEA.  
 DR GO: GO:0004812; F:RNA ligase activity; IEA.  
 DR GO: GO:0006418; P:amino acid activation; IEA.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR008957; FN\_III-LIKE.  
 DR InterPro: IPR007110; IG-LIKE.  
 DR InterPro: IPR003598; IG\_c2.

DR InterPro: IPR000048; IQ\_region.  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR000219; RHGEF.  
 DR InterPro: IPR001452; SH3.  
 DR InterPro: IPR001412; RNA-eync\_I.  
 DR Pfam: PF00041; fn3; 2.  
 DR Pfam: PF00047; Ig; 49.  
 DR Pfam: PF00612; IQ; 1.  
 DR Pfam: PF00169; PH; 1.  
 DR Pfam: PF00621; RHGEF; 1.  
 DR SMART: SM00060; FN3; 2.  
 DR SMART: SM00408; IGC2; 17.  
 DR SMART: SM00015; IQ; 1.  
 DR SMART: SM00233; PH; 1.  
 DR SMART: SM00326; SH3; 1.  
 DR PROSITE: PS00178; AA\_TRNA\_LIGASE\_I; 1.  
 DR PROSITE: PS50010; DH\_2; 1.  
 DR PROSITE: PS50835; IG\_LIKE; 46.  
 DR PROSITE: PS50096; IQ; 1.  
 DR PROSITE: PS50003; PH\_DOMAIN; 1.  
 KM Immunoglobulin domain.  
 SO SEQUENCE 6620 AA; 721665 MW; C2A8BB77B284452 CRC64;

Query Match 5.7%; Score 133.5; DB 4; Length 6620;  
 Best Local Similarity 22.5%; Pred. No. 2.9;  
 Matches 91; Conservative 59; Mismatches 154; Indels 101; Gaps 20;

DR 34 GDTVELICTASQKSIQFHKNKSNQIKILGNQSFILTKGSKINDPADRRS-LMD--Q 89  
 DB 2841 GEDVELRCEIS-RAGTPVHM-----LKDRAIKRSQKYVCE 2877  
 QY 90 GNFP-LIIKNIKIEDSDTYICEVEDQKEVOLLVPLGLTANSPTHL-----QGSL-T 140  
 DB 2878 GTYAMLVIRKASIKDGEYTCVEASKSTASLHV-EKANCFEELTNLQVEKGTAVF 2936  
 QY 141 LTLESPPGSSPVQCR-----SPRGKNIQCGKT--LSVQLELDQSGTWCTVLQNO 190  
 DB 2937 CKTEHP---AATYTWKGLLELRASGKHQPSQEGLLRLTISLKEKADSTTYCDIGQAQ 2993  
 QY 191 KVEPKI-----DIVLAFQKASSIVYKKEGQVEFPFLAFYVEK----- 231  
 DB 2994 SRAQLLVGRRVHIIDLEVDVQEGSSATFRCRISPAVNE-FVHFPLDKTPHLANELNE 3052  
 QY 232 -----LFGSGELMWQ--ERRASSKSWITFDLKNKEVSVKRVTDPKL 272  
 DB 3053 IDAQPGYHVLTLRQALADSGTYFEAGDQASAA-----LWTEKPSVFSFELTDAT 3107  
 QY 273 QMGKLLPLHL-TLQALPYAGSGNLTLEAKTGKLHQ--EYNLVMBAT-QLQKNLTG 328  
 DB 3108 TEGEDTLVCEITCQIPMCWTYDGTILGSAACQLSHGHRAQLITGATTLQDSGRYK 3167  
 QY 329 EVMGPTSPKIMLSKLENKREKVPVVLNPAQNMOCILS 373  
 DB 3168 EAGGACSSSI---VRVHARPVRFQALKDIEVLGGAATLRCVLS 3209

RESULT 65  
 Q8ISF3 PRELIMINARY; PRT; 2693 AA.  
 AC Q8ISF3;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE 301kDa\_2 protein.  
 GN ISOF.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematozoa; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=2269627; PubMed=12381307;  
 RA Flaherty D., Gernert K., Shmeleva N., Tang X., Mercer K.,

RA Borodovsky M., Benian G.;  
 RT "Titins in Caenorhabditis elegans with Unusual Features: Coiled-coil  
 RT Domains, Novel Regulation of Kinase Activity and Two New Possible  
 RT Elastic Regions";  
 RL J. Mol. Biol. 323:533-549(2002).  
 DR EMBL: AY130758; AAN61521.1; -.  
 DR GO: GO:0005524; F:ATP binding; IEA.  
 DR GO: GO:0004674; F:protein serine/threonine kinase activity; IEA.  
 DR GO: GO:0004713; F:protein-tyrosine kinase activity; IEA.  
 DR GO: GO:0004648; P:protein amino acid phosphorylation; IEA.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR007110; IG\_1like.  
 DR InterPro: IPR003598; IG\_c2.  
 DR InterPro: IPR000719; Prot\_kinase.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.  
 DR InterPro: IPR008271; Ser\_thr\_kin\_AS.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR Pfam: PF00041; fn3; 2.  
 DR Pfam: PF00047; ig; 19.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Prodom: PD000001; Prot\_kinase; 1.  
 DR SMART: SM00060; FN3; 1\_kinase; 1.  
 DR SMART: SM00408; IGC2; 19.  
 DR SMART: SM00220; S\_TKC; 1.  
 DR SMART: SM00219; TYKC; 1.  
 DR PROSITE: PS50835; IG\_LIKE; 18.  
 DR PROSITE: PS50107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR SEQUENCE 2693 AA; 299094 MW; 1C64AD2011B8391A CRC64;

Query Match 5.7%; Score 133; DB 5; Length 2693;  
 Best Local Similarity 20.7%; Pred. No. 0.92;  
 Matches 94; Conservative 68; Mismatches 161; Indels 132; Gaps 22;  
 QY 28 VILGKGGDTVELICTASQKSIQFHWKNSNQIKLGNQGSFLTQKPSKLNDRADRRSLW 87  
 DB 1100 VILKTAGEFATFTCOSYANPAQVW-----LHNGKALQOTKSNYKTRLF 1144  
 QY 88 DQGNFPLIIKMLKIDSDTYICEVEDQKEV-----QLVFGILTANS----- 129  
 DB 1145 DQNTATLVLENVTDLCGTTAVANNQFGVHTSAQLTISGSEAKKIASLPHYITELKP 1204  
 QY 130 DTHLQGSGLTLTLSPGSSP-----SVQCRSPR-----GKNIQSGKTLVSQ 173  
 DB 1205 KINWVG--ATLSIQADLNGSPIPEVWMLKDNSELVESDRIMKCDGVVYQ----LLVND 1258  
 QY 174 LELQDSGTWCTVLOÑQKVEFKIDIVLAFOKASSIYKKEGEQVEF-----S 222  
 DB 1259 VGLDEBGTVITIAENEKGIHQNTQEVSV--TKSKEVEKEKKEKKEKDEGKKRGRPG 1315  
 QY 223 FPL---AFTVEKLT-----GSGELMWQAEASASSKSWTF-DLKNKEVSVK----- 264  
 DB 1316 LPRPSGASKTEQVYMAFADSEGPADSYEVERRCPDQREWVSCGSTKSLELEIKGILTPT 1375  
 QY 265 ----RVTQDPKLOMGKKLPLHLTLPOA---LPQYAGSGNLTLALBAKTGKLAHQEVNLV 316  
 DB 1376 EYIFRVAGKNKQGLGEMSEMTSLTKTASVGQAQF-----TISPSK-----ILA 1420  
 QY 317 MRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREK-----PWWVLN----PEA 365  
 DB 1421 NRDDPEF--IAVEFSGTPRP-----SVKMKENQIIVDEKIDIVATTSTSIILMLKQEN 1474  
 QY 366 GMMQCLL-SDSGQVLEESN-----KVLPTWSTEVH 395  
 DB 1475 GTFNCLLENLQGSASCOVITNKPASLQSTPDH 1509

RESULT 66  
 O8ISF4 PRELIMINARY; PRT; 2708 AA.  
 AC O8ISF4;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE 301kDa\_1 protein.  
 GN ISOR.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nemtoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OC NCBI\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22269627; PubMed=12381307;  
 RA Flaherty D., Gernert K., Shmeleva N., Tang X., Mercer K.,  
 RA Borodovsky M., Benian G.;  
 RT "Titins in Caenorhabditis elegans with Unusual Features: Coiled-coil  
 RT Domains, Novel Regulation of Kinase Activity and Two New Possible  
 RT Elastic Regions";  
 RL J. Mol. Biol. 323:533-549(2002).  
 DR EMBL: AY130758; AAN61520.1; -.  
 DR GO: GO:0005524; F:ATP binding; IEA.  
 DR GO: GO:0004674; F:protein serine/threonine kinase activity; IEA.  
 DR GO: GO:0004713; F:protein-tyrosine kinase activity; IEA.  
 DR GO: GO:0004648; P:protein amino acid phosphorylation; IEA.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR007110; IG\_1like.  
 DR InterPro: IPR003598; IG\_c2.  
 DR InterPro: IPR000719; Prot\_kinase.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.  
 DR InterPro: IPR008271; Ser\_thr\_kin\_AS.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR Pfam: PF00041; fn3; 2.  
 DR Pfam: PF00047; ig; 19.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Prodom: PD000001; Prot\_kinase; 1.  
 DR SMART: SM00060; FN3; 1\_kinase; 1.  
 DR SMART: SM00408; IGC2; 19.  
 DR SMART: SM00220; S\_TKC; 1.  
 DR SMART: SM00219; TYKC; 1.  
 DR PROSITE: PS50835; IG\_LIKE; 18.  
 DR PROSITE: PS50107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR SEQUENCE 2708 AA; 300729 MW; 2BBEBB856FA4571 CRC64;

Query Match 5.7%; Score 133; DB 5; Length 2708;  
 Best Local Similarity 20.7%; Pred. No. 0.92;  
 Matches 94; Conservative 68; Mismatches 161; Indels 133; Gaps 22;  
 QY 28 VILGKGGDTVELICTASQKSIQFHWKNSNQIKLGNQGSFLTQKPSKLNDRADRRSLW 87  
 DB 1115 VILKTAGEFATFTCOSYANPAQVW-----LHNGKALQOTKSNYKTRLF 1159  
 QY 88 DQGNFPLIIKMLKIDSDTYICEVEDQKEV-----QLVFGILTANS----- 129  
 DB 1160 DQNTATLVLENVTDLCGTTAVANNQFGDVHTSAQLTISGSEAKKIASLPHYITELKP 1219  
 QY 130 DTHLQGSGLTLTLSPGSSP-----SVQCRSPR-----GKNIQSGKTLVSQ 173  
 DB 1220 KINWVG--ATLSIQADLNGSPIPEVWMLKDNSELVESDRIMKCDGVVYQ----LLVND 1273  
 QY 174 LELQDSGTWCTVLOÑQKVEFKIDIVLAFOKASSIYKKEGEQVEF-----S 222  
 DB 1274 VGLDEBGTVITIAENEKGIHQNTQEVSV--TKSKEVEKEKKEKKEKDEGKKRGRPG 1330  
 QY 223 FPL---AFTVEKLT-----GSGELMWQAEASASSKSWTF-DLKNKEVSVK----- 264  
 DB 1331 LPRPSGASKTEQVYMAFADSEGPADSYEVERRCPDQREWVSCGSTKSLELEIKGILTPT 1390  
 QY 265 ----RVTQDPKLOMGKKLPLHLTLPOA---LPQYAGSGNLTLALBAKTGKLAHQEVNLV 316  
 DB 1391 EYIFRVAGKNKQGLGEMSEMTSLTKTASVGQAQF-----TISPSK-----ILA 1435  
 QY 317 MRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREK-----PWWVLN----PEA 365

Db 1436 NRDEFE--IAVEFSGTPPT----SVKWKENIQIVDEKIDVATTSTSIILNKSQEN 1489  
QY 366 GMMQCLL-SDSGQVLEESNI-----KVLPTWSTPH 395  
Db 1490 GTFNCLLENELGQASASCOVTFINPKPASLOSTPDH 1524

RESULT 67  
Q8MNS0 PRELIMINARY; PRT; 2780 AA.  
ID Q8MNS0  
AC Q8MNS0;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN F12F3.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxId=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA Waterston R.;  
RT "genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Fulton B., Mohlmann P.;  
RT "The sequence of C. elegans cosmid F12F3.";  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U80022; AAM29672.1; -.  
DR WormPep; F12F3.2a; CE30753.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.  
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.  
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
DR InterPro; IPR003962; FNIII subd.  
DR InterPro; IPR003961; FN\_III-like.  
DR InterPro; IPR007110; IG\_1like.  
DR InterPro; IPR003598; IG\_c2.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_kinase.  
DR InterPro; IPR008271; Ser\_thr\_kin\_AS.  
DR InterPro; IPR001245; Tyr\_kinase.  
DR Pfam; PF00041; fn3; 2.  
DR Pfam; PF00047; ig; 19.  
DR Pfam; PF00069; pkinase; 1.  
DR PRINTS; PRO0014; FNTYPEIII.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00060; FN3; 2.  
DR SMART; SM00408; IGc2; 19.  
DR SMART; SM00220; S\_TKc; 1.  
DR SMART; SM00219; TYKc; 1.  
DR PROSITE; PS50835; IG\_LIKE; 18.  
DR PROSITE; PS50107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR Hypothetical protein; ATP-binding; Immunoglobulin domain; Repeat;  
KM transfease.  
SQ SEQUENCE 2780 AA; 308646 MW; 046D057107B935C1 CRC64;

Query Match 5.7%; Score 133; DB 5; Length 2780;

Best Local Similarity 20.7%; Pred. No. 0.96;  
Matches 94; Conservative 68; Mismatches 161; Indels 132; Gaps 22;  
QY 28 VILGKKGDTVELTCTASQKKSIOFHWKNSNQIKILNQSFLLTKGSKINDRADRSRLW 87  
Db 1187 VVILKTAGEFATFCYGSANPAQVW-----LHNGALQQTCKSNYKTRLF 1231  
QY 88 DQGNPFLIINKIKIEPSDPTICEVEPOKEEV-----OLVFGLTANS----- 129  
Db 1232 DQNTATLVLENVYDELGGTYTAVANNQFGDVHTSAQLTISGSEAKTIASLPFIIELEK 1291  
QY 130 DTHLGOSILVLTLESPPSSP-----SVQCRSP-----GKNIOGKTLVSQ 173  
Db 1292 KINVVG--ATLSIQDLNMGSPPEVWMLKDNSELVESRIDIQKCGGVYQ-----LLVND 1345  
QY 174 LEIQDSGTWTCYVLQNKVYEFKIDIVLAFOKASSIVYKKGEOVF-----S 222  
Db 1346 VGLDEGTGYTTIAENKGRIRQTEVSV--TSKSVKKEKKEKKYKDEKKKGRBQ 1402  
QY 223 FPL---AFTVEKLT-----GSGLMWAERASSSKSMITP-DLKNKEVSVK----- 264  
Db 1403 LPRPSGASTBOVTMAFPAPBGPADSYVERKCPQREWVSGSTKSLELEIKGLTPNT 1462  
QY 265 ---RYTODPKLQMGKKLPLHLTPQA---LPQVAGSNTLTALBAKTGKLHQEVNLVY 316  
Db 1463 EYIFRAGNKKQGLGEMSEMTSLKTAASVGAPOF-----TISPSK-----ITA 1507  
QY 317 MRATQLOKULTCGWCPTSPKMLSLKLENKAKVSKREK-----PWWVLN-----PEA 365  
Db 1508 NRDEFE--IAVEFSGTPPT----SVKWKENIQIVDEKIDVATTSTSIILNKSQEN 1561  
QY 366 GMMQCLL-SDSGQVLEESNI-----KVLPTWSTPH 395  
Db 1562 GTFNCLLENELGQASASCOVTFINPKPASLOSTPDH 1596

RESULT 68  
Q8MNS1 PRELIMINARY; PRT; 2808 AA.  
ID Q8MNS1  
AC Q8MNS1;  
DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN F12F3.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxId=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA Waterston R.;  
RT "genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium.";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Fulton B., Mohlmann P.;  
RT "The sequence of C. elegans cosmid F12F3.";  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Bristol N2;  
RA Waterston R.;  
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U80022; AAM29673.1; -.  
DR WormPep; F12F3.2b; CE30754.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.  
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.

DR GO:0016740, F:transferase activity, IEA.  
 DR GO:0006468, P:protein amino acid phosphorylation, IEA.  
 DR InterPro: IPR003963, FN III.  
 DR InterPro: IPR003961, FN III.  
 DR InterPro: IPR007110, Ig-like.  
 DR InterPro: IPR003598, Ig\_c2.  
 DR InterPro: IPR000719, Prot kinase.  
 DR InterPro: IPR002290, Ser thr kinase.  
 DR InterPro: IPR008271, Ser thr pkin\_AS.  
 DR InterPro: IPR001245, Tyr\_kinase.  
 DR Pfam: PF00041, fn3\_2.  
 DR Pfam: PF00047, Ig\_19.  
 DR Pfam: PF00069, pkinase\_1.  
 DR PRINTS: PR00014, FNTYPEIII.  
 DR ProDom: PD000001, 2\_kinase\_1.  
 DR SMART: SM00060, FN3\_2\_kinase\_1.  
 DR SMART: SM00408, IGC2\_19.  
 DR SMART: SM00220, S\_TKC\_1.  
 DR SMART: SM00219, TYKC\_1.  
 DR PROSITE: PS00835, IG LIKE\_18.  
 DR PROSITE: PS00107, PROTEIN\_KINASE\_ATP\_1.  
 DR PROSITE: PS0011, PROTEIN\_KINASE\_DOM\_1.  
 DR PROSITE: PS00108, PROTEIN\_KINASE\_ST\_1.  
 DR Hypothetical protein; ATP-binding; Immunoglobulin domain; Repeat; transferase.  
 KW TRANSFERENCE.  
 SQ SEQUENCE 2808 AA; 311979 MM; 02AFED0AFE06FE12 CRC64;

Query Match 5.7%; Score 133; DB 5; Length 2808;

Best Local Similarity 20.7%; Pred. No. 0.97; Matches 94; Conservative 68; Mismatches 161; Indels 132; Gaps 22;

DR 28 VVLGKKGDTVELTCTASQKSIQFHWKNSNQIKILNGQSFLLTKPSKINDRADSRSLM 87  
 DB 1187 VVLKTAGEYATPTCCSYANPAQVW-----LHNGKALQOTKSNYKTRLF 1231  
 QY 88 DQGNFPLIKNLKIEDSDTYICEVEDEKEV---QLVFGLTANS----- 129  
 DB 1232 DDMTATLVIENTDELCTGTAVANNQFGDVHTSAQLTISGEBAKKIASLPHYIIELPK 1291  
 QY 130 DTHLQGGSLTLTLSPSSP-----SVQCRSPR-----GKNIGGKTLVSQ 173  
 DB 1292 KINWVG--ATLSIQADLNGSPIPEVWMLKONSELVESDRIQMKCDGVNYQ---LLVND 1345  
 QY 174 LELQDSGTWTCYVLQONQKVEFKIDIVVLAFOKASSIVYKKEGEQVEF-----S 222  
 DB 1346 VGLDEGTYYTITTAENEKGIKIRONTESV---TSKVEKKEKKYKKEKDEGKKRGRPG 1402  
 QY 223 FPL---AFVVEKLT-----GSGELMWQAEPAASSKSWITP-DLKNKEVSVK----- 264  
 DB 1403 LPRPSGASKTEQYTMAFDAPSEGPADSYEVERRCPPQREWVSCGSTKSLLEIKGLTPNT 1462  
 QY 265 ----RYTQDPKLOMGKKLPLHLTLPOA---LPOYAGSGNLTLALAKTGKHLQEVNLV 316  
 DB 1463 EYIFRPAKGNKKGGESEMTSTLKTASVGAPOF-----TISPSK----- 1507  
 QY 317 MRATOLQKNLTCVWGPTSPKMLSLLENKEAKVSKREK-----PVVLN-----PRA 365  
 DB 1508 NRNDDEP--IAVEFSGPTP-----SVMKYENLQIYVDEKIDVATSTSTSLNLKSGEEN 1561  
 QY 366 GMMQCLL-SDSGVLLLESNI-----KVLPTWSTPVH 395  
 DB 1562 GTENCLEIENELQGSASQCVTIFNKPASLQSTPDH 1596

RESULT 69

Q8ISF6 PRELIMINARY; PRT; 18519 AA.  
 AC Q8ISF6;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE 2MDA\_2 protein.  
 GN ISOF.

OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 ON NCBI\_TaxID=6239;  
 RX MEDLINE=2269627; PubMed=12381307;  
 RA Flaherty D., Gernert K., Shmeleva N., Tang X., Mercer K.,  
 RA Bordovsky M., Benlian G.;  
 RT "Titins in Caenorhabditis elegans with Unusual Features: Coiled-coil  
 RT Domains, Novel Regulation of Kinase Activity and Two New Possible  
 RT Elastic Regions.";  
 RL J. Mol. Biol. 323:533-549(2002).  
 DR EMBL; AY130758; AAM61518.1; -.  
 DR PIR; E89066; E89066.  
 DR PIR; T33247; T33247.  
 DR GO:0005524, F:ATP binding, IEA.  
 DR GO:0008237, F:metalloproteinase activity, IEA.  
 DR GO:0004674, F:protein serine/threonine kinase activity, IEA.  
 DR GO:0004713, F:protein-tyrosine kinase activity, IEA.  
 DR GO:0008270, F:zinc ion binding, IEA.  
 DR GO:0006468, P:protein amino acid phosphorylation, IEA.  
 DR GO:0006508, P:proteolysis and peptidolysis, IEA.  
 DR InterPro: IPR008938, ARM.  
 DR InterPro: IPR001064, Crystalin.  
 DR InterPro: IPR003961, FN III.  
 DR InterPro: IPR008957, FN III-like.  
 DR InterPro: IPR001023, Hsp70.  
 DR InterPro: IPR007110, Ig-like.  
 DR InterPro: IPR003598, Ig\_c2.  
 DR InterPro: IPR003596, Ig\_v.  
 DR InterPro: IPR006025, Pept\_M\_Zn\_BS.  
 DR InterPro: IPR000719, Prot kinase.  
 DR InterPro: IPR002290, Ser thr kinase.  
 DR InterPro: IPR008271, Ser thr pkin\_AS.  
 DR InterPro: IPR001245, Tyr\_kinase.  
 DR Pfam: PF00041, fn3\_11.  
 DR Pfam: PF00047, Ig\_43.  
 DR Pfam: PF00069, pkinase\_1.  
 DR ProDom: PD000089, Hsp70\_1.  
 DR ProDom: PD000001, Prot\_kinase\_1.  
 DR SMART: SM00060, FN3\_11.  
 DR SMART: SM00408, IGC2\_37.  
 DR SMART: SM00406, IGV\_4.  
 DR SMART: SM00220, S\_TKC\_1.  
 DR SMART: SM00219, TYKC\_1.  
 DR PROSITE: PS00225, CRYSTALLIN\_BETAGAMMA\_1.  
 DR PROSITE: PS00835, IG\_LIKE\_38.  
 DR PROSITE: PS00107, PROTEIN\_KINASE\_ATP\_1.  
 DR PROSITE: PS00011, PROTEIN\_KINASE\_DOM\_1.  
 DR PROSITE: PS00108, PROTEIN\_KINASE\_ST\_1.  
 DR PROSITE: PS00142, ZINC\_PROTEASE\_1.  
 SQ SEQUENCE 18519 AA; 2051869 MM; 6A8441C5D0BA7729 CRC64;

Query Match 5.7%; Score 133; DB 5; Length 18519;

Best Local Similarity 20.7%; Pred. No. 13; Matches 94; Conservative 68; Mismatches 161; Indels 132; Gaps 22;

DR 28 VVLGKKGDTVELTCTASQKSIQFHWKNSNQIKILNGQSFLLTKPSKINDRADSRSLM 87  
 DB 16926 VVLKTAGEYATPTCCSYANPAQVW-----LHNGKALQOTKSNYKTRLF 16970  
 QY 88 DQGNFPLIKNLKIEDSDTYICEVEDEKEV---QLVFGLTANS----- 129  
 DB 16971 DDMTATLVIENTDELCTGTAVANNQFGDVHTSAQLTISGEBAKKIASLPHYIIELPK 17030  
 QY 130 DTHLQGGSLTLTLSPSSP-----SVQCRSPR-----GKNIGGKTLVSQ 173  
 DB 17031 KINWVG--ATLSIQADLNGSPIPEVWMLKONSELVESDRIQMKCDGVNYQ---LLVND 17084  
 QY 174 LELQDSGTWTCYVLQONQKVEFKIDIVVLAFOKASSIVYKKEGEQVEF-----S 222  
 DB 17085 VGLDEGTYYTITTAENEKGIKIRONTESV---TSKVEKKEKKYKKEKDEGKKRGRPG 17141

```
QY 223 FPL---AFTVEKLT-----GSGELMWQAERASSSKSWITP-DLKNEVSVK----- 264
DB 17142 LPRPSGASKEQVYTMFADAPSEGPADSYEVERRCPPQREWVSGSGTSLELEIKGLTPNT 17201
QY 265 ----RVTQDPKLOMGKKLPLHLTLPPQ----LPQVAGSGNLTALAEKTKGLHQEVNLV 316
DB 17202 EYIFRVAGKKNKQGLGEMSEMTSTLKTASVGAPOF-----TISPSK-----IIA 17246
QY 317 MRATQLOKMLTCEVWGPTSPKMLSLKLENKAVSKREK-----PVMVLN----PEA 365
DB 17247 NRDDFEF--IAVEFSGTPP-----SVKMYKENLIQVDEKIDVATTSTSSILMLKSGEEN 17300
QY 366 GMMQCLL-SDSGVLLNESNI-----KVLPTWSTPVH 395
DB 17301 GTFNCLLENELGQASASCQVTFENKPASLQSTPDH 17335

RESULT 70
Q8ISF7 PRELIMINARY; PRT; 18534 AA.
AC Q8ISF7;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE 2MDA_1 protein.
GN ISOF.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pseudocercariae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22269627; Pubmed=12381307;
RA Flaherty D., Gernert K., Shmeleva N., Tang X., Mercer K.,
RA Borodovsky M., Benian G.;
RA "Fluorescence resonance energy transfer (FRET) analysis of the
RT domain, Novel Regulation of Kinase Activity and Two New Possible
RT Elastic Regions."
RL J. Mol. Biol. 323:533-549(2002).
DB EMBL; AY130758; ANA61517.1; -.
DR PIR; E89066; E89066.
DR PIR; T33247; T33247.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0004674; F:Protein serine/threonine kinase activity; IEA.
DR GO; GO:0004713; F:Protein-tyrosine kinase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR001064; Cysteallin.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR008957; FN III-like.
DR InterPro; IPR001023; Hep70.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR003596; IG_v.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_Thr_kinase.
DR InterPro; IPR008271; Ser_Thr_kin_AS.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00041; fn3; 11.
DR Pfam; PF00047; ig; 43.
DR Pfam; PF00069; Pkinase; 1.
DR Prodom; PD000089; Hep70; 1.
DR SMART; SM000001; Prot_kinase; 1.
DR SMART; SM00060; fn3; 11.
DR SMART; SM00408; IGC2; 37.
DR SMART; SM00406; IGV; 4.
DR SMART; SM00220; STKC; 1.
DR SMART; SM00219; TyKC; 1.
```

```
DR PROSITE; PS00225; CRYSTALLIN BETAGAMMA; 1.
DR PROSITE; PS50835; IG_LIKE; 38.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ SEQUENCE 18534 AA; 2053504 MW; BC6A682B943C8C0A CRC64;

Query Match 5.7%; Score 133; DB 5; Length 18534;
Best Local Similarity 20.7%; Pred. No. 14;
Matches 94; Conservative 68; Mismatches 161; Indels 132; Gaps 22;

DB 28 VVIGKGDVVELCTGASQKSIQFHKNSNQIKILNQSGSFLTKGSKLNDRAISRSM 87
16941 VVKTGEGTATFTCCGYANPAQVW-----LHNGALQOTGSNTYTRLF 16985
QY 88 DQGNFPLIKMLKIEDSDTYICEVEDQKEV-----QLVFGLTANS----- 129
DB 16986 DQNTATLVLENVTDELGGTYTAVANNQPGDVHTSAQLTISGSAKKIASLPYFIELKP 17045
QY 130 DTHLOGSLTLTLESPPSSP-----SVQCRSR-----GKNIQSKTISVQ 173
DB 17046 KINWVG--ATLSIQADLNGSPPEVWMLKDNSELVESDRIQMKDGVNYQ----LLVAD 17099
QY 174 LELDSGVTWCTVLOKQKVEFKIDIVLAFQKASIVYKKEGEYF-----S 222
DB 17100 VGLDEGTYITTDENKGIKQNTVEVS--TKSKVEKKEKKKVEKDEGKKRG 17156
QY 223 FPL---AFTVEKLT-----GSGELMWQAERASSSKSWITP-DLKNEVSVK----- 264
DB 17157 LPRPSGASKEQVYTMFADAPSEGPADSYEVERRCPPQREWVSGSGTSLELEIKGLTPNT 17216
QY 265 ----RVTQDPKLOMGKKLPLHLTLPPQ----LPQVAGSGNLTALAEKTKGLHQEVNLV 316
DB 17217 EYIFRVAGKKNKQGLGEMSEMTSTLKTASVGAPOF-----TISPSK-----IIA 17261
QY 317 MRATQLOKMLTCEVWGPTSPKMLSLKLENKAVSKREK-----PVMVLN----PEA 365
DB 17262 NRDDFEF--IAVEFSGTPP-----SVKMYKENLIQVDEKIDVATTSTSSILMLKSGEEN 17315
QY 366 GMMQCLL-SDSGVLLNESNI-----KVLPTWSTPVH 395
DB 17316 GTFNCLLENELGQASASCQVTFENKPASLQSTPDH 17350

RESULT 71
Q9W6V2 PRELIMINARY; PRT; 352 AA.
AC Q9W6V2;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Neurocractin-L.
GN NTRA-L.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Brummenndorf T.;
RA TISSUE=Brain;
RA Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99264333; Pubmed=10330412;
RA Marg A., Strim P., Spaltmann F., Plagge A., Kauselmann G., Buck F.,
RA Rathjen F.G., Brummenndorf T.;
RA "Neurocractin, A novel neurite outgrowth-promoting Ig-like protein
RT that interacts with CPU-1 and LAMP."
RL J. Cell Biol. 145:865-876(1999).
```

DR EMBL: AJ132999; CAB4446.1; -.  
 DR InterPro: IPR007110; IG\_1like.  
 DR InterPro: IPR003598; IG\_c2.  
 DR Pfam: PF00047; Ig; 3.  
 DR SMART: SM00408; IGG2; 2.  
 DR PROSITE: PS0835; IG\_LIKE; 3.  
 DR Immunoglobulin domain.  
 SQ SEQUENCE 352 AA; 37944 MW; B0FA99F295FD8FA0 CRC64;

Query Match 5.7%; Score 132.5; DB 13; Length 352;  
 Best Local Similarity 21.9%; Pred. No. 0.059;

Matches 76; Conservative 55; Mismatches 111; Indels 105; Gaps 17;

QY 20 PAATQGNKVVLTGKGGDTVELTCTASQKSIQFHMKNSNOIKILGNGSFLTKGSPKLANDR 79  
 DB 39 PGAAADSLVY--RKGTAVLRCLYLEDGAS--KGAMLRSSIIIFAG-----SDK 82  
 QY 80 --ADSRSL--WDQGNFPLIKKLIKEDSDTYICEVEDQKE-----EVQLLVFGL 125  
 DB 83 WSVDPVSIATANRREYSIQIDVDYTDGPTCSQTOHTPTMQVHLTVQVSPKIFRI 142  
 QY 126 TANSDFHLLQGSGLTTLSPSSPSVQCR--SPRGKNIQGGKTLVSQGLQDSGTWT 183  
 DB 143 --SSDIVNBSGVTLVCLATGKPEPSISMRHISPSAKPPESQYLDIGITRDQGEYE 200  
 QY 184 CTVLQKQKVEFKIDIVLAFQKASSIVYKKEGEQYEFPLAFTVEKLTGSGEL----- 238  
 DB 201 CS-----AENDVSVPDVKKVKNVNFAPITQELKSSGVMGNG 239  
 QY 239 -----WQNER--ASSKSWITFDLKNKE--VSVKRTQD----- 269  
 DB 240 LIRCEGAGVPAVPEWYRGERKLISGQOGIT--IKNVSTRSLTVNVTNVEHFGNYCVA 297  
 QY 270 -PRLQW--GKKLPLHLTPQALPOYAGSN-----LTALAEKT 305  
 DB 298 AKKLGMNNSLPLN---PESTAQYGTIGDAEVLFCMWLVLTLSLT 341

## RESULT 72

Q9HCL6 PRELIMINARY; PRT; 1596 AA.

AC Q9HCL6;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein KIAA1556 (Fragment).  
 GN KIAA1556.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20450683; PubMed=10997877;  
 RA Nagase T., Kikuno R., Nakayama M., Hirose M., Ohara O.,  
 RT "Prediction of the coding sequences of unidentified human genes.  
 RT XVII. The complete sequences of 100 new cDNA clones from brain which  
 RT code for large proteins in vitro."  
 RL DNA Res. 7:273-281(2000).  
 DR EMBL: AB046776; BAB13382.1; -.  
 DR InterPro: IPR007110; IG\_1like.  
 DR Pfam: PF00047; Ig; 17.  
 DR SMART: SM00408; IGG2; 7.  
 DR PROSITE: PS0835; IG\_LIKE; 15.  
 KM Hypothetical protein; Immunoglobulin domain.  
 FT NON TER 1  
 FT NON TER 1596  
 SQ SEQUENCE 1596 AA; 175280 MW; 88937E0760C4E2EF CRC64;

Query Match 5.7%; Score 132.5; DB 4; Length 1596;  
 Best Local Similarity 21.2%; Pred. No. 0.48;  
 Matches 90; Conservative 59; Mismatches 136; Indels 139; Gaps 20;

QY 34 GDTVELTCTASQKSIQFHMKNSNOIKILGNGSFLTKGSPKLANDRARSRS-LND---Q 89  
 DB 116 GEDVELRELRS--RACTPVHM-----LKRKAIRKSKQKDVYCE 152  
 QY 90 GNPF-LIIRKLIKEDSDTYICEVEDQKEVOLVFGTLTANSPTHL-----QGSL-T 140  
 DB 153 GTMALVIRGASLKDAGEYCEVEASKSTASLHV--EEXANCTBELTINLOVEKGTAVPT 211  
 QY 141 LTLESPPSSPSVQCR-----SPRGKNIQGGKTLVSQGLQDSGTWTCTVLQNG 190  
 DB 212 CTEHP---AATVWRKGLLEIRASGKQPSQEGTLRLTTSALKABSDDTYICDQAG 268  
 QY 191 KKEFKI-----DIVLAFQKASSIVYKKEGEQYEFPLAFTVEK----- 231  
 DB 269 SNAQLLVGGRVHIHLEDVDVQDGSATPRCIRSPANYE--PVHMFDPKPLHANELNE 327  
 QY 232 -----LTSGELMWQ--ERASSKSWITFDLKNKEVSVKRTQDPRK 272  
 DB 338 IDAQPGYHVLTRQALAKDSGTIFYEAGDGRASAA-----LRVTEKPSV 372  
 QY 273 QMKKLPPLHLTPQALPOYAGSNLTALAEKT-----GKL-----H 309  
 DB 373 FSRBLTDATTEGE-----DLTLVCESTCDIPVCMTKQGTTLGSAQCQLSHSGH 423  
 QY 310 QEVNLVWRATQQLQKNLTCVWGPSTPKLMLSLKENKAKVSKREKPVVNLNPEAGMWQ 369  
 DB 424 RAQLLITGATQIDQSGRYCEAGACSSSI---VRHAPVRQGLMKOLEVEGGAATLR 480  
 QY 370 CLLS 373  
 DB 481 CVLS 484

## RESULT 73

Q9VNP2 PRELIMINARY; PRT; 811 AA.

AC Q9VNP2;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE CG8779 protein.  
 GN NRM OR CG8779.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Suton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Fiankoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu U., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borova K.A., Botchan M.R., Bouck J., Brokstein P., Broctier P.,  
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaeser K.,  
 RA Glodek A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.U., Wei M.-H., Ibegwam C.,  
 RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,



RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mitsuhashi N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,  
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Rehner K., Remington K., Saunders R.D.C., Scheeler P., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spter E., Spreading A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirbas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*,"  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Krommiller B., Man K.H., Holt R.A.,  
 RA Evans C.A., Gockyne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doug L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ileguez C., Jalili M., Kruse D., Li P., Matei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacle J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Peltman G.S., Puri V., Richarde S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskae R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,  
 RT "Sequencing of *Drosophila melanogaster* genome,"  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Miara S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Bernier B., Carlson J.W., Celniker S.E.,  
 RA Clamp W., Drysdale R., Emert D., Frise E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,  
 RT "Annotation of *Drosophila melanogaster* genome,"  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.,  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AEO03599; AAF51887.2; -.  
 DR FlyBase: FBgn0005629; nrm.  
 DR GO: GO:0005886; C:plasma membrane; IDA.  
 DR InterPro: IPR002086; Alddehyde\_dehydrt.  
 DR InterPro: IPR003599; IG.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003598; IG\_c2.  
 DR Pfam: PF00047; IG\_6.  
 DR SMART: SM00409; IG\_5.  
 DR SMART: SM00408; IGc2; 4.  
 DR PROSITE: PS00070; ALDEHYDE\_DEHYDR\_CYS; 1.  
 DR PROSITE: PS50835; IG\_LIKE; 7.  
 KM Immunoglobulin domain.  
 SQ SEQUENCE 811 AA; 90442 MW; BAC689EA2C8E23F4 CRC64;  
 Query Match 5.7%; Score 132; DB 5; Length 811;  
 Best Local Similarity 21.1%; Pred. No. 0.2;  
 Matches 83; Conservative 68; Mismatches 141; Indels 102; Gaps 20;  
 QY 10 LLLVQLALLPAAT-OGNKVVLGKGGDTVELTCTASQKSIQPH---WKNSQNI--KIL 62

Db 26 LVVLVLTALVDSSTAQVDTTISQGESQSVLP-PCPVDAEKCGKLSLNMFKGDRIRIAML 85  
 QY 63 GNOGSFLTGPSPKLNDRADSRSLMDQGNPPLIKLKLTEDSTTYICEVEDQKEEVOLV 122  
 Db 86 GD-----SNVTSVNFKEPERVTV-EQNPRYLVIKDLKIADEIDYLCDT-----T 128  
 QY 123 FGITANSDFHLLOGSGLTLTLSPSPSSPVQCRSRGNIGCGKLTSLVSQLELDPSGWT 182  
 Db 129 FFIPESTCDN-FNGYRIELRLVLP--TEVILDAKGRIRIKGSV--VGPMDEROSLKA 182  
 QY 183 TCTVLONOKKVEKIDIVLAFQKASIVYKKEGSEVSPFLPAFTVKL-TGSGELMWQ 241  
 Db 183 TCTVNRTRPQPE-----VSMFGTKRLTTPSPHLDVGLYTSLELDWT 227  
 QY 242 AEPASSSKSWITFDLKNKEVSVKRTQ---DPLQD-----GKULPLH 282  
 Db 228 LSRREDLAOD--TECRVKSAAIQNTVTRFSYDLQVPTSIDNGVKHHTVQSGSKVLTLC 284  
 QY 283 TLFPQALPOV-----AGSGNLT---LALFAKTKLHQEVLNVMRAOTLOXN- 325  
 Db 285 DIFGAPPAVNLWYNTTTTISGENEITVRSKSLKSDGTFHTQSEL-IFNAITRENDR 343  
 QY 326 -LTCEVWGPTSPKMLSLKLENKAKVSKREKPV 358  
 Db 344 VERCEN-----ENIVLQIN-REKRI 362  
 RESULT 74  
 Q24273 PRELIMINARY; PRT; 1011 AA.  
 AC 024273;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)  
 DE NEUROMUSCULIN.  
 GN NRM OR CG8779.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_Taxid=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94000831; PubMed=8398154;  
 RA Kania A., Han P.L., Kim Y.T., Belien H.,  
 RT "Neuromusculin, a *Drosophila* gene expressed in peripheral neuronal  
 RT precursors and muscles, encodes a cell adhesion molecule,"  
 RL Neuron 11:673-687(1993).  
 DR EMBL: L23146; AAA03750.1; -.  
 DR PIR: T13669; T13669.  
 DR FlyBase: FBgn0005629; nrm.  
 DR GO: GO:0005886; C:plasma membrane; IDA.  
 DR InterPro: IPR003599; IG.  
 DR InterPro: IPR007110; IG-like.  
 DR Pfam: PF00047; IG\_8.  
 DR SMART: SM00409; IG\_5.  
 DR PROSITE: PS50835; IG\_LIKE; 9.  
 KM Cell adhesion.  
 SQ SEQUENCE 1011 AA; 113125 MW; AFD8A2A015D3AE63 CRC64;  
 Query Match 5.7%; Score 132; DB 5; Length 1011;  
 Best Local Similarity 21.1%; Pred. No. 0.28;  
 Matches 83; Conservative 68; Mismatches 141; Indels 102; Gaps 20;  
 QY 10 LLLVQLALLPAAT-OGNKVVLGKGGDTVELTCTASQKSIQPH---WKNSQNI--KIL 62  
 Db 26 LVVLVLTALVDSSTAQVDTTISQGESQSVLP-PCPVDAEKCGKLSLNMFKGDRIRIAML 85  
 QY 63 GNOGSFLTGPSPKLNDRADSRSLMDQGNPPLIKLKLTEDSTTYICEVEDQKEEVOLV 122  
 Db 86 GD-----SNVTSVNFKEPERVTV-EQNPRYLVIKDLKIADEIDYLCDT-----T 128

QY 123 FGLTANS DTHLLOGSITLTLESPPGSPVOCSPRGKNIQGGKTLVSQLELQDSGTW 182  
DB 129 FFIPEFTCON-FNGYRIELRLVLP---TEVILDAKGRIKNGSV--VGPMEQSGSLKA 182  
QY 183 TCTVLONOKKVEERKIDIVLAFQKASSIYKKEGEVRSFPALFVEKL-TGSGELMWQ 241  
DB 183 TCTVTRRTRPQPE-----VSWFRGTKRLLTYSPYTHDLVDGLYSTLELDWT 227  
QY 242 AERASSKSMITFDLKNKEYSVKRYTQ--DPLQOM-----GKRLPLHL 282  
DB 228 LSHEDLADQ---IECRVKSALIONVYTKFSVDLQVRPFSIDINGVKHHTVQSGKVVLYC 284  
QY 283 TLFPQALPOY-----AGSGNLT---LALBAKTGKLHQEVNLVYMRATOLQXN- 325  
DB 285 DINGAPPAVNLTYWNTTTIISGENEITEVRSKLESDQTFHTQSEL-IFNATRENDR 343  
QY 326 -LTCEVWGPTSPKMLSLKLENKEAKVSKREKV 358  
DB 344 VFRCEA-----ENIVLDIN-REKPI 362

## RESULT 75

Q9XT41 PRELIMINARY; PRT; 1248 AA.  
ID Q9XT41  
AC Q9XT41;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Neural cell adhesion molecule L1.  
GN L1CAM.  
OS Cercopithecus aethiops (Green monkey) (Grivet).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecinae; Cercopithecus.  
NCBI\_TaxID=9534;  
RX MEDLINE=20259241; PubMed=10797421;  
RA Finch U., Schroder J., Resaler B., Veeke A., Gal A.;  
RT "Spectrum and detection rate of L1CAM mutations in isolated and  
familiar cases with clinically suspected Li-disease.";  
RL Am. J. Med. Genet. 92:40-46(2000).  
DR EMBL; AF129167; AAD28610.1; -.  
DR HSSP; P20241; ICFB.  
DR InterPro; IPR003961; FN\_III.  
DR InterPro; IPR008957; FN\_III-like.  
DR InterPro; IPR007110; I9-like.  
DR InterPro; IPR003598; I9\_c2.  
DR Pfam; PF00041; fn3; 5.  
DR Pfam; PF00047; I9; 6.  
DR SMART; SM00060; FN3; 4.  
DR SMART; SM00408; IGC2; 5.  
DR PROSITE; PSS0835; IG\_LIKE; 6.  
KW Immunoglobulin domain.  
SQ SEQUENCE 1248 AA; 138901 MW; 2438A93A29C3BD61 CRC64;

Query Match 5.7%; Score 132; DB 6; Length 1248;  
Best Local Similarity 19.4%; Pred. No. 0.37;  
Matches 99; Conservative 66; Mismatches 191; Indels 130; Gaps 21;  
QY 8 RHLLLV-----LQALLPA--ATGKNKVVLGKKGDTVELTCTASQKKSIOFHWKNSNQIKI 61  
DB 404 RHGLLLANAYIVVQLPKILITADNQTVMVAVGSAVILCKAFGAPVPSVQMLDEGDTTV 463  
QY 62 LGNGSFLTGKPSKLANDRARSRLMDQGNFLLIKNLKIEBDDTYICEVEDQKEBYOLL 121  
DB 464 LQDERFP-----PYANGTLGIRDLRANDTGRYFCLAAANDQNNV-TI 503  
QY 122 VFGLTANS DTHLLO-----GQSLTFLTES--PPGSSPVOCSPRGKNIQ----- 164  
DB 504 VAHLKYKDATQITQGGPRSAIEKKGSRVITTCQASFPDPSLOPSTIRKGD-GRDLQELGDS 562

QY 165 -----GKTLVSQLELQDSGTWITCTVLQNOKKVEFKIDIVLAF-----QKASSIYKK 214  
DB 563 KYFIEDGR-LVIHSLDSDQGNYSVASTELDVESRQQLLVGSPGPVRLVLDLHL 621  
QY 215 EGEVFEFSPPLA-----FTVEKLTSGSELIMQAEARASSSKSNITPDLNKEYSV 263  
DB 622 TQSOVRVSWSPAEHDNAPIEKYDIEFEDKEKMAPKMYSLGKVPQSTTLTLKLSPYVHT 681  
QY 264 KRVTDPLQMGKPLHLTL--POLPO-----YAGSGNLTLEAKTGKLHQEVNLVY 316  
DB 682 FRYVTLINKYGRGEPSPVSETVTPPALAPKPNVDYKGEENET-----TNMYI 728  
QY 317 MRATOLQKNLTCEVGP-----TSPKMLSLKLENKEAKVSKREKVPVNLNPEAGMQ- 369  
DB 729 -----TWKPLRMMDMNAPOVYVRVQ-----WRPQGTGRGPMQE 760  
QY 370 CLLSGQVLESNIKVLPTMSTPYHPAPASALPAPPTGSALPDPQTASALPPPPAASALP 429  
DB 761 QIVSD--PFLVVSNTSTP---VPEIRVQAVNSQKG--PEPQVTTIGSGEDYPQALP 811

## RESULT 76

Q7YQ17 PRELIMINARY; PRT; 1255 AA.  
ID Q7YQ17  
AC Q7YQ17;  
DT 01-OCT-2003 (TREMBLrel. 25, Created)  
DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE L1 cell adhesion molecule (Fragment).  
GN L1CAM.  
OS Pongo pygmaeus (Orangutan).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Pongo.  
NCBI\_TaxID=9600;  
RX MEDLINE=22763540; PubMed=12777533;  
RA Kitano T., Schwarz C., Nickel B., Paabo S.;  
RT "Gene Diversity Patterns at 10 X-Chromosomal Loci in Humans and  
Chimpanzees.";  
RL Mol. Biol. Evol. 20:1281-1289(2003).  
DR EMBL; AB102655; BAC81124.1; -.  
FT NON\_TER 1  
SQ SEQUENCE 1255 AA; 139782 MW; 4FDAFCDB12629C0 CRC64;

Query Match 5.7%; Score 132; DB 6; Length 1255;  
Best Local Similarity 19.4%; Pred. No. 0.38;  
Matches 93; Conservative 65; Mismatches 192; Indels 130; Gaps 21;  
QY 8 RHLLLV-----LQALLPA--ATGKNKVVLGKKGDTVELTCTASQKKSIOFHWKNSNQIKI 61  
DB 407 RHGLLLANAYIVVQLPKILITADNQTVMVAVGSAVILCKAFGAPVPSVQMLDEGDTTV 466  
QY 62 LGNGSFLTGKPSKLANDRARSRLMDQGNFLLIKNLKIEBDDTYICEVEDQKEBYOLL 121  
DB 467 LQDERFP-----PYANGTLGIRDLQANDTGRYFCLAAANDQNNV-TI 506  
QY 122 VFGLTANS DTHLLO-----GQSLTFLTES--PPGSSPVOCSPRGKNIQ----- 164  
DB 507 VANLKYKDATQITQGGPRSAIEKKGSRVITTCQASFPDPSLOPSTIRKGD-GRDLQELGDS 565  
QY 165 -----GKTLVSQLELQDSGTWITCTVLQNOKKVEFKIDIVLAF-----QKASSIYKK 214  
DB 566 KYFIEDGR-LVIHSLDSDQGNYSVASTELDVESRQQLLVGSPGPVRLVLDLHL 624  
QY 215 EGEVFEFSPPLA-----FTVEKLTSGSELIMQAEARASSSKSNITPDLNKEYSV 263  
DB 625 TQSOVRVSWSPAEHDNAPIEKYDIEFEDKEKMAPKMYSLGKVPQSTTLTLKLSPYVHT 684  
QY 264 KRVTDPLQMGKPLHLTL--POLPO-----YAGSGNLTLEAKTGKLHQEVNLVY 316  
DB 685 FRYVTLINKYGRGEPSPVSETVTPPALAPKPNVDYKGEENET-----TNMYI 731

Oy	317	HRAQTOLQNLNCEVWGP-----TSPKMLSLKLENKEAKNSKRREPVWVLNPEAGMNO-	365
Dd	732	-----TKPRLRMWDNAPVOYRRQO-----WRPGSTGRPWQE	763
Oy	370	CLLSDSGVLLLESNIKVLPTWSPPHNPASALPAEPGTSALPDPOGTASALPDPNPASALP	429
Dd	764	QIVSD--PFLWMSNSTSF----VPRIKIQVANNSGKG---PEPVTIGYSGEDIPQALP	814
 RESULT 77			
Q8NHN3		PRELIMINARY;	PRT; 2212 AA.
ID	Q8NHN3		
AC	Q8NHN3:		
DT	01-OCT-2002 (TREMBLrel. 22, Created)		
DT	01-OCT-2002 (TREMBLrel. 22, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE	Obscurin (Fragment).		
GN	OBSCN.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Mammalia; Eutereria; Primates; Carnivora; Hominiidae; Homo.		
OK	NCBI_TaxId=9606;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RA	Young P., Ehler E., Gautel M.;		
RT	"Obscurin, a giant sarcomeric Rho-GEF protein involved in sarcomere		
RT	assembly.";		
RL	Submitted (May-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AJ314905; CACB5752.1; -.		
DR	InterPro; IPR003599; IG_.		
DR	InterPro; IPR007110; IG-like.		
DR	InterPro; IPR003598; IG_c2.		
DR	Pfam; PF00047; Ig_24.		
DR	SMART; SM00409; IG_25.		
DR	SMART; SM00408; IGc2; 23.		
DR	PROSITE; PS50835; IG_LIKE; 22.		
KW	immunoglobulin domain.		
FT	NON_TER 1		
FT	NON_TER 1		
SQ	SEQUENCE 2212 AA; 241980 MW; B6DFA263EB749465 CRC64;		
 Query Match 5.7%; Score 132; DB 4; Length 2212; Best Local Similarity 21.2%; Pred. No. 0.83; Matches 90; Conservative 59; Mismatches 136; Indels 140; Gaps 20;			
Oy	34	GDTVELLTASQSKSIQFHMKNSNQIKILNGSFLTGTSPKLNLRADSRSS-LMD---Q	89
Dd	108	GEDVELRELSS-RAGTPVM-----LKQRKAIRKSQKDVACE	144
Oy	90	GNFP-LITIKLIKIEDSDTYICEVDQKEVOLVFGLTANSPTHL-----QGOSL-T	140
Dd	145	GTMAMLVIRGASLKQAGEYTCVEASKSTASHV-EKANCTTEELTLQLQVEEKTAFT	203
Oy	141	LTLSEPPSSPSVSOCR-----SPRGNIQGGKT--LSVSOLELDGSGTWTCVLQNQ	190
Dd	204	CKTEHP---AATVTRKGLLELRASGKHQPSEGTLTLRILTALKADSDTYTCIDIGAQ	260
Oy	191	KKVERKI-----DIVLAFQASSIYVKGEQVESFPLAFYER-----	231
Dd	261	SRAQLLVQAARRVHIIEDEDVDVGESSATFRCSISPANE-PVHWFLDKLPRLHANELN	319
Oy	212	-----LTGSSELMOQA--ERASSSKWITFDLKNKEVSVKRVTDQPK	271
Dd	320	EIDAQPGGVHVLTQLQLAKDSGTLYPEAGDGRASA-----LARTKPS	364
Oy	272	LQMGKKLPLHLTPLPALFOYAGSGULTALEAKT-----GKL-----	308
Dd	365	VFSRELDTATTITEG-----DLTLVCETSTDIPVCWKDKRTLRSARCOLSHHG	415
Oy	309	HOEVLVVMMRATOLQNLNCEVWGTSPTLMSLTENKEAVSSRKRPVWVLNPEAGMM	368
Dd	416	HRAQLLTIGATLQDGRKCEGAGCASSI---VAVNAFPVRFGALDKLEVLEGAAFL	472

QY	369	QCLHS	373	
Db	473	RCVLS	477	
RESULT 78				
ID	Q23550	PRELIMINARY;	PRT,	6831 AA.
AC	Q23550	Q23020; Q2723;		
DT	01-NOV-1996	(TEMBLrel. 01, Created)		
DT	01-NOV-1998	(TEMBLrel. 08, Last sequence update)		
DT	01-OCT-2003	(TEMBLrel. 25, Last annotation update)		
DE	UNC-22 protein.			
GN	UNC-22 OR ZK617.1.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea			
OC	Rhabditidae; Pelodierinae; Caenorhabditis.			
OX	NCBI_TaxId=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	White S., Harris B.;			
RL	Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 784-6831 FROM N.A.			
RC	STRAIN=BRISTOL N2;			
RX	MEDLINE=90044042; PubMed=2812002;			
RA	Benian G.M., Kiff J.E., Neckelmann N., Moerman D.G., Waterson R.H.;			
RT	"Sequence of an unusually large protein implicated in regulation of			
RL	myosin activity in C. elegans.";			
RL	Nature 342:45-50(1989).			
RN	[3]			
RP	SEQUENCE OF 784-6831 FROM N.A.			
RC	STRAIN=BRISTOL N2;			
RX	MEDLINE=93387664; PubMed=8397135;			
RA	Benian G.M., l'Hernault S.W., Morris M.E.;			
RT	"Additional sequence complexity in the muscle gene, unc-22, and its			
RL	encoded protein, twitchin, of Caenorhabditis elegans.";			
CC	-1- ALTERNATIVE PRODUCTS:			
CC	Event=Alternative splicing; Named isoforms=2;			
CC	Name=a;			
CC	IsoId=Q23550-1; Sequence=Displayed;			
CC	Note=No experimental confirmation available;			
CC	Name=b;			
CC	IsoId=Q23551-1; Sequence=External;			
CC	Note=No experimental confirmation available;			
CC	-1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.			
DR	EMBL; Z73897; CAA98064.1; JOINED.			
DR	EMBL; Z73899; CAA98081.1; -			
DR	EMBL; Z73897; CAA98081.1; JOINED.			
DR	EMBL; X15423; CAA33463.1; -			
DR	PIR; A88852; A88852.			
DR	PIR; S57242; S57242.			
DR	HSSP; Q63450; IA06.			
DR	GO; GO:0005524; F:ATP binding; IEA.			
DR	GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.			
DR	GO; GO:0016740; F:proteinase activity; IEA.			
DR	GO; GO:0004812; F:RNA ligase activity; IEA.			
DR	GO; GO:0006418; P:amino acid activation; IEA.			
DR	GO; GO:0006468; P:protein amino acid phosphorylation; IEA.			
DR	InterPro; IPR003962; FNIII subd.			
DR	InterPro; IPR003964; FN.III.			
DR	InterPro; IPR008957; FN.III-like.			
DR	InterPro; IPR007110; Ig-like.			
DR	InterPro; IPR003598; Ig_C2.			
DR	InterPro; IPR000719; Prot_kinase.			
DR	InterPro; IPR002290; Ser_Thr_kinase.			
DR	InterPro; IPR008271; Ser_Thr_kin_AS.			
DR	InterPro; IPR001412; tRNA-synt_1.			
DR	Pfam; PF00041; fn3_31.			
DR	Pfam; PF00047; Ig_21.			
DR	Pfam; PF00069; pkinase; 1.			



```

Qy 225 LAFTVEKLTGSGELMWAERASSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPLHLTL 284
Db 1519 FQI---KGRRGD-----PKQILLKNGKPIDEENRKL-VEYITIKDVAIEVFK----- 1562
Qy 285 PQALPQVAGSGNLTALAEAKTGKLEHVNLMV-----RATOLQKNLTGE---VWGPT 334
Db 1563 ---NPQALADTGKMLBELGNSAGTALAPFELFVKDKPKPKSPLETNKVTAEGDLVWGTP 1619
Qy 335 SPKMLSLK--LENKE-----AKVSK-----REKPVVNLPEAGMQLSLD 374
Db 1620 DPEGAGVKKYIIMQEGRSGGNMAKVGETKGTDFKVDKDLKEHGEKERVXA-LNECGLSD 1678
Qy 375 --SGQVLESNIKVLPTMSPTVHPR 397
Db 1679 PLTGESVLAKN-----PIYGVGKPK 1698

RESULT 80
ID 025198 PRELIMINARY; PRT; 848 AA.
AC 025198;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Receptor tyrosine kinase 90.
GN HTRK90.
OS Hydra attenuata (Hydra) (Hydra vulgaris).
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Anthomedusae;
OC Hydrae; Hydra.
OX NCBI_TaxId=6087;
RN [1]
RP SEQUENCE FROM N.A.
RA Miller M.A., Steele R.E.;
RT "The Hydra Receptor Tyrosine Kinase, HTRK90, is Expressed in a Subset
of the Interstitial Cell Population.";
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U59448; AAB03389.1; -.
DR HSSP: P08631; IAD5.
DR GO: GO:0005524; P:ATP binding; IEA.
DR GO: GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003598; IG_c2.
DR InterPro: IPR007119; Prot_kinase.
DR InterPro: IPR001245; Tyr_kinase.
DR InterPro: IPR008266; Tyr_kinase_AS.
DR Pfam: PF00047; IG; 5.
DR Pfam: PF00069; Pkinase; 1.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Prot_kinase; 1.
DR SMART: SM00408; IGC2; 2.
DR SMART: SM00219; TYRC; 1.
DR PROSITE: PS50835; IG LIKE; 5.
DR PROSITE: PS50107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS50011; PROTEIN KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN KINASE_TYR; 1.
KM ATP-binding; Immunoglobulin domain; Kinase; Transferase;
KM Tyrosine-protein kinase.
SQ SEQUENCE 848 AA; 95551 MW; DOA52ED6A8760C07 CRC64;

Query Match 5.7%; Score 131.5; DB 5; Length 848;
Best Local Similarity 20.0%; Pred. No. 0.24;
Matches 87; Conservative 63; Mismatches 181; Indels 103; Gaps 18;

Qy 37 VELTCTSOAKSIQFHMKNISQIKILNGSFLTKGSKLNDRAD-SRRSLMDQGNPLI 95
Db 48 VFVNCCTSPVEDAQTII-----NKGEENIPSYSEGTKEVYVLPNNTLY 91
Qy 96 IKNLKIEDSDTYICEVDQKEEV-QLLVF-----GLTANSPTHLLOGOSLTTLLESPPG 148
Db 92 LKSLKKKEAGTYPCNAFLTEKLFQTVVIEVAFLOKLNISIDPVKVLGVYVAIEICEPVG 151

```

```

Qy 149 S-SPSYQ-----CRSPRGKNIQGGK-TLSVSOLEIDSGTWTCTVLQN--OKKVEPKID 198
Db 152 KPQPKKMLYNGATITSPDNPNDISSWTLRIKTKMLSDKQNYCVAPNSAETPAERKLT 211
Qy 199 IVVIAFOKASSIYK-KEGQVFEFSF-----PLAFTVEKLTGSGELMWAERASSKSW 251
Db 212 VVVVGVDVPSIENSIKEGSATFKCESKSDPL-----VYKMRDLETGSTT- 259
Qy 252 ITFDLKNKEVSVKRVTPDPLQMGKKLPLHLTLPQALPQVAGSGNLTALAEAKTGKLEH 311
Db 260 -----DNSQVCKYKEDVIPSDRVYISGGTISIRNAVKQEGIYLCETELSTQISAK 313
Qy 312 VNLVWRARATOLQKNL-----TCEVWGPTSPKMLSLKLENKEKVSREKPVVNL----- 362
Db 314 VQLNVELMKVDKSLNLETQCDRLRQNNTKTICHERFGGKYSIQWTRLSAPNALSNMK 373
Qy 363 -----PEAGMQC-----LSDSGQVLE-SNIKVL---PTW 390
Db 374 VVNDISYISLKEKEDMGQPCFCAVGEYNNATAVNFIYVEHQPIISPKNITAYIGEPMW 433
Qy 391 STVPHPASALPAP 404
Db 434 ---VHCQGGKFPKP 444

RESULT 81
ID 07SX76 PRELIMINARY; PRT; 483 AA.
AC 07SX76;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE FGF receptor-like protein precursor (Hypothetical protein).
GN FGFR1L.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxId=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Trub B., Zhuang L., Taeschler S., Wiedemann M.;
RT "Characterization of FGFR1L, a novel FGF receptor preferentially
expressed in skeletal tissues.";
RL J. Biol. Chem. 278:0-0(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=22386257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haish F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA Krzyzanski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;

```



```
QY 122 VFGLTANSDFHLQ-----GSLTLTLES--PPGSSPVQCRSPRGKNIQ----- 164
DB 508 A-NLKVDATQITQGPSTTEKKSRYFTCOASFPDLSLQSTWRGD--GRDQLQELQSD 565
QY 165 -----GCKTISVSOLELQDSGTWTCTVLQOKKVEFKIDIVLAF-----QKASIIYKK 214
DB 566 KYFIEDGR-LVHSLDYSDQGNVSCVASTELDVESAOQLLVGSPGPRVLVLSDLHL 624
QY 215 EGEVESSFPLA-----FTVEKLTGSGELMQAERASSKSWITFDLKNKEVS 263
DB 625 TQSGVRSWSPADHDNAPIEKYDIEFEDKEMAPEKWSLQKVPQNGSTTLTKLSPVYHT 684
QY 264 KRYVQDPKQMGKKLPLHLTL--POALPQ-----YAGSGNLTALAEAKTKLHQEVNLV 316
DB 685 FRVTAINKYGPSPSVSETVTPPEAPENPMDVKGBNET-----TNWVI 731
QY 317 MRATOLQKNLTCEVWGP-----TSPKMLSLKLENKAKVSKREKPVWLINPEAGMWQ- 369
DB 732 -----TWKPLRMWMDNAPQVQYRVQ-----WRPGQTRGPWQE 763
QY 370 CLISDSGOVLLESNIKVLPTWSTFVHPRASALPAPPTGSALPDPQTASALPDPPEASALP 429
DB 764 QIVSD--PFLVVSNTSTF---VPEIKVQAVNSQCKG---PEPQVITIGYSGEDYPOAIP 814

RESULT 84
QY0L8 PRELIMINARY; PRT; 1255 AA.
AC QY0L8;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE L1 cell adhesion molecule (Fragment).
GN L1CAM.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
OX NCBI_TaxID=9598;
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE=22763540; PubMed=12777533;
RA Kltano T., Schwarz C., Nickel B., Paabo S.;
RT "Gene Diversity Patterns at 10 X-Chromosomal Loci in Humans and
RT Chimpanzees.";
RL M01. Biol. Evol. 20:1281-1289(2003).
DR EMBL; AB102654; BAC81123.1; -.
FT NON TER 1
SQ SEQUENCE 1255 AA; 139742 MW; 235D47715A3DD6D9 CRC64;

Query Match 5.6%; Score 131; DB 6; Length 1255;
Best Local Similarity 19.4%; Pred. No. 0.45;
Matches 93; Conservative 65; Mismatches 192; Indels 130; Gaps 21;

QY 8 RHLLLV-----LQALPAP--ATGQNKVVLGKGGDTVELTCTASQKKSIOFHMKNNSNQIKI 61
DB 407 RHGILLNNAVITYVQQLPAKILLADNQYMAVQSTAYLLCKAGAPVPVQMIDEDGTTV 466
QY 62 LGNQSFLTKGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEQQL 121
DB 467 LQDERFP-----PYANGTLGIRLQANDICRYCLAAANQNNV-TI 506
QY 122 VFGLTANSDFHLQ-----GSLTLTLES--PPGSSPVQCRSPRGKNIQ----- 164
DB 507 VANLKVDATQITQGPSPAIEKKSGRYFTCOASFPDLSLQSTWRGD--GRDQLQELQSD 565
QY 165 -----GCKTISVSOLELQDSGTWTCTVLQOKKVEFKIDIVLAF-----QKASIIYKK 214
DB 566 KYFIEDGR-LVHSLDYSDQGNVSCVASTELDVESAOQLLVGSPGPRVLVLSDLHL 624
QY 215 EGEVESSFPLA-----FTVEKLTGSGELMQAERASSKSWITFDLKNKEVS 263
DB 625 TQSGVRSWSPADHDNAPIEKYDIEFEDKEMAPEKWSLQKVPQNGSTTLTKLSPVYHT 684
```

```
QY 264 KRYVQDPKQMGKKLPLHLTL--POALPQ-----YAGSGNLTALAEAKTKLHQEVNLV 316
DB 685 FRVTAINKYGPSPSVSETVTPPEAPENPMDVKGBNET-----TNWVI 731
QY 317 MRATOLQKNLTCEVWGP-----TSPKMLSLKLENKAKVSKREKPVWLINPEAGMWQ- 369
DB 732 -----TWKPLRMWMDNAPQVQYRVQ-----WRPGQTRGPWQE 763
QY 370 CLISDSGOVLLESNIKVLPTWSTFVHPRASALPAPPTGSALPDPQTASALPDPPEASALP 429
DB 764 QIVSD--PFLVVSNTSTF---VPEIKVQAVNSQCKG---PEPQVITIGYSGEDYPOAIP 814

RESULT 85
QY0N2 PRELIMINARY; PRT; 1340 AA.
AC QY0N2;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
GN DKFPZ434P0216.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA TISSUE=Testis;
RA Koehler K., Beyer A., Mewes H.W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL834139; CAD38854.1; -.
DR InterPro; IPR007110; I9-1like.
DR Pfam; PF00047; I9_15.
DR SMART; SM00408; IGc2; 14.
DR PROSITE; PS50835; IG_LIKE; 14.
KW Hypothetical protein; Immunoglobulin domain.
FT NON TER 1
SQ SEQUENCE 1340 AA; 143889 MW; 7AC26BE86DB7FBF CRC64;

Query Match 5.6%; Score 131; DB 4; Length 1340;
Best Local Similarity 19.7%; Pred. No. 0.49;
Matches 80; Conservative 58; Mismatches 163; Indels 106; Gaps 16;

QY 12 LVQLALLPRAATGQNKVVLGKGGDTVELTCTASQKKSIOFHMKNNSNQIKILGNQSFLLK 71
DB 929 LVVQVPE--FVIEGRLDLSLTBGSNAFLPCKARGSPENITMDKQQ--PVSGAEKFTT- 984
QY 72 GPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVED---QKEVQLLVFGL-- 125
DB 985 -----QSGELVNLLEGODAGTYTCAENAVGARARRVHLTIIVLPV 1027
QY 126 --TANSDFHLQGSLLTLTLESPPGSSPVQCR--SPRGKNI--QGCKTISVQLEIQ 177
DB 1028 FTLVPCDRSLRGLDRMLCAAGSPTRIGMTVNDPRTEGSEBDGSGSTLQRAAVSRH 1087
QY 178 DSGTWTCTVLQOKKVEFKIDIVLAFQKASSIVYKKEGEVQEFSPPLAFTEVKLTGS- 235
DB 1088 DSGTTCYCA--ENR-----VGRITQAVSFVHKAPVLC--GEAFYIVLPEVGSIG 1134
QY 236 -----GELMQAERASSKSWITFDLKNKEVSVTRVTQDPKQMGKKLPLHLTL 285
DB 1135 LDCVWAGDPVPRDIHWIKDILPRGSHLRQQLONGSLTIHRTDR----- 1178
QY 286 QALPQVAGSGNLTALAEAKTKLHQEVNLVWRAATQLOKN-----LTGEVWG 332
DB 1179 -----DGRVYQCLAENEMGAKKVILVLQSAFVQVEPQDMYTRSGDDVALRCQATG 1231
QY 333 PTPSKMLSLKLENKAKVSKR-----EKPVWLINPE--AGMWQCT 371
DB 1232 EPTP--TIEMLOAGPRLRASRRRLTLPDGSIMLENVETQDAGTYDCV 1276
```





DR	GO:	GO:0005509;	F:calcium ion binding; IEA.
DR	GO:	GO:0005215;	F:transporter activity; IEA.
DR	GO:	GO:0006810;	P:transport; IEA.
DR	InterPro:	IPR000152;	Axh_hydroxyl_1_S.
DR	InterPro:	IPR000515;	BPD_transp.
DR	InterPro:	IPR001861;	EGF_Ca.
DR	InterPro:	IPR006209;	EGF_Like.
DR	InterPro:	IPR007110;	IG_Like.
DR	InterPro:	IPR003598;	IG_C2.
DR	InterPro:	IPR002035;	VWF_A.
DR	Pfam:	PF00047;	IG: 47.
DR	SMART:	SM00179;	EGF_CA; 1.
DR	SMART:	SM00408;	IGC2: 44.
DR	SMART:	SM00327;	VMA; 1.
DR	PROSITE:	PS00010;	ASX_HYDROXYL; 1.
DR	PROSITE:	PS00402;	BPD_TRANSF_INN_MEMBER; 1.
DR	PROSITE:	PS01186;	EGF_2; 1.
DR	PROSITE:	PS01187;	EGF_CA; 2.
DR	PROSITE:	PS50835;	IG_LIKE; 47.
KW	EGF-like domain;		Immunoglobulin domain; Signal.
FT	SIGNAL	1	POTENTIAL.
FT	CHAIN	25	5198
SO	SEQUENCE	5198	AA; 570809
			KW; DA8511F2B58D37B
			CRC64;
			HEMICENTIN.

[illegible]

CC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Breast muscle;  
RX MEDLINE=96254045; PubMed=8660363;  
RA Yajima H., Ohtsuka H., Kawamura Y., Kume H., Murayama T., Abe H.,  
RA Kimura S., Maruyama K.;  
RT "A 11.5-kb 5'-terminal cDNA sequence of chicken breast muscle  
RT connectin/rutin reveals its Z line binding region.";  
RL Biochem. Biophys. Res. Commun. 223:160-164(1996).  
RN [2]  
RN SEQUENCE OF 1767-1871 FROM N.A.  
RX MEDLINE=96365546; PubMed=8769723;  
RA Tuncaloglu K.T., Mittal B., Sanger J.W., Sanger J.W.;  
RT "Partial characterization of zeugmatin indicates that it is part of  
RT the Z-band region of rutin.";  
RL Cell Motil. Cytoskeleton 34:108-121(1996).  
CC -I- SIMILARITY: CONTRAINS 33 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.  
DR EMBL; D83390; BA011908.1; -.  
DR EMBL; U64829; AAC60019.1; -.  
DR PIR; T42633; T42633.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003598; IG\_C2.  
DR InterPro; IPR003006; IG\_MHC.  
DR Pfam; PF00047; Ig; 33.  
DR SMART; SM00408; IGC2; 14.  
DR PROSITE; PSS0835; IG\_LIKE; 29.  
DR PROSITE; PSS0290; IG\_MHC; 1.  
DR Immunoglobulin domain.  
FT KW  
FT SEQUENCE 4162 AA; 464971 MW; DB0905C907686649 CRC64;  
SQ

Query Match	5.6%	Score 130;	DB 13;	Length 4162;
Best Local Similarity	22.4%	Pred. No. 2.9;		
Matches 103;	Conservative 55;	Mismatches 159;	Indels 142;	Gaps 24;
QY	11	LTVIOLMLPPAATGKGVKGGDPTEVLTCTASQKKSIOFHW-KNSNOI-----KILG	63	
		: : :       : : :       : : :       : : :		
Db	3551	LTVIEPAATVEKPGPVKT---AGDSCTLECTVDGTPELTARWFRDGNELSTDHKYKI--	3605	
		: : :       : : :       : : :       : : :		
QY	64	NOGSFLAT--GPKSLKNDRADRSRLMPDQGNPELLIKNLIKEDSDTYICEVDOKEEVOLL	121	
		: : :       : : :       : : :       : : :		
Db	3606	---SFFNKVSGSLKILN-----AGLEBDSGETFEFVKNSVSGKSSCTASIQVSDR-----	3649	
		: : :       : : :       : : :       : : :		
QY	122	VFGITANSDFHLLQGSLLTLE--SPGSSPSVQCC--SP-----RGKNIQGGK	167	
		: : :       : : :       : : :       : : :		
Db	3650	-----IMPPSFRKLKETYQGLGSSAVILECKVYSGPILVSWFHFDGQETSGD	3697	
		: : :       : : :       : : :       : : :		
QY	168	-----TLSVSOLELDOSGTWTCTVIONKVEFKIDIVLAFQKASSIYYKKE-	215	
		: : :       : : :       : : :       : : :		
Db	3698	KYQATLTLDNTGSLKVNGLQESDMGTYSTCTAIVNAGSDCS---AFLSVREPPSPFVKPEP	3754	
		: : :       : : :       : : :       : : :		
QY	216	-----GEVYESSFLATFVEKLTGSGELMW---QAEKRSASSKMITPPLKNKEVSVKVT	267	
		: : :       : : :       : : :       : : :		
Db	3755	FNVLSGENITFT-----SIVKSGSPPLEKVMRGSIELAPGHKNITL-----	3796	
		: : :       : : :       : : :       : : :		
QY	268	QD--PKLOMGKKLPLHLTPQALPQVYAGSGLLTALBAKTKLQEVNL-----VYMR	318	
		: : :       : : :       : : :       : : :		
Db	3797	QDSVAELELEFVQVPLQ-----SGDYTCQVSNKACKISCTTHLFLFKPEAKFVWK	3844	
		: : :       : : :       : : :       : : :		
QY	319	ATOLO---KNLTCEVMGPTSPKMLSLKLNKEAKVSKREKPVVNLNPEA-----	365	
		: : :       : : :       : : :       : : :		
Db	3845	VNDLSVEKGNLILECTYTGTTPPISVYWK---KNGVILKSHKKSITTTETSAILEIPNS	3901	
		: : :       : : :       : : :       : : :		
QY	366	-----GMMQC--LLSDSGVLLSNIKVL--PTWSTPVHP	396	
		: : :       : : :       : : :       : : :		
Db	3902	KLSDQGYVSGHIENDSGQDNCHGAILLEPPYFVTLPEP	3940	
		: : :       : : :       : : :       : : :		
RESULT	89			
ID	088654			
	088654			
PRELIMINARY:				
	PRT: 340 AA.			

```

AC 088554;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DN NK cell receptor 284 splice variant.
GN NMRK.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA Stepp S.E., Schatzle J.D., Bennett M., Kumar V., Mathew P.A.;
RT "Characterization of genomic structure and alternative splicing of the
RT murine NK cell receptor 284."
RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF082803; AAC34859.1; -.
DR MGD; MGI:109294; Nmrk.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR SMART; SM00409; IG_1.
DR KMR.
SQ SEQUENCE 340 AA; 38397 MW; 0D5275A1A354FC1E CRC64;

Query Match 5.6%; Score 129.5; DB 11; Length 340;
Best Local Similarity 23.7%; Pred. No. 0.095;
Matches 72; Conservative 50; Mismatches 121; Indels 61; Gaps 11;

QY 18 LRLPATQG-----NRVLGKGDYELTCTASQKSIQFHWKNSNQ-----IKILGNQG 66
DB 12 LLRLAHQGDQPSSEEVGVSGKPVQLRPSNIQTKDVSVQWKTGSHRKIEIL----- 67
QY 67 SFLTKGPSKLNDRADSRSLMDQGNPFLIKNLIKIEDSDTYICEVDQKEV-----QLL 121
DB 68 NWNNDGPSWNSVPSDIYG-FDYGDPALSTKSAKLQDSGHYLLIINTGKVKCNKPFOLL 126
QY 122 VFGLTANSDTHLQGG-----SLTTLTSPGSSPVQCRSPRGKNI-----QGGKTLAV 171
DB 127 ILD---HVERPNLKAQWKPTNGTCQLFLSCLVTKDQNVYALRGSTLISNRNSTHWE 183
QY 172 SQLELDQSGTGT-----TVLONKRVFKIDIVVLAFOKASST- 210
DB 184 NQIDASLSLHTYTCNVSRASMANHTLNFTHGQSVSNFRFLPFYIVILVTLFLGAIL 243
QY 211 ---VYKKEGQVPSPLATVEKLTGSGELMWQAE---RASSSKSWITFDLKNKEVSVK 264
DB 244 CFCVWTKRKQQLQFSPKEPLTIYVYKDSRPSRDQGCSPSAPVQEDGRGQRELD 303
QY 265 RVTQ 268
DB 304 RVSE 307

RESULT 90
Q9D7B8 PRELIMINARY; PRT; 287 AA.
AC Q9D7B8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 2310016B05RIK protein.
GN D11ERTD736E OR 2310016B05RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Tongue;
RA MEDLINE=2108566; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

```

```

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H.,
RA Kuenli P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudil F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bulc C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Williams L.,
RA Wymshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL; AK009375; BAB26251.1; -.
DR MGD; MGI:1289168; D11ERTD736E.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR Pfam; PF00047; IG_1.
DR SMART; SM00409; IG_1.
DR PROSITE; PS50835; IG-LIKE; 1.
SQ SEQUENCE 287 AA; 32061 MW; D932AD3308A03639 CRC64;

Query Match 5.5%; Score 129; DB 11; Length 287;
Best Local Similarity 25.2%; Pred. No. 0.082;
Matches 54; Conservative 36; Mismatches 84; Indels 40; Gaps 9;

QY 8 RLHLVLQLALPPA--ATQGNKVLGKGDYELTCTASQK-XSIOFHWKNSNQIK----- 60
DB 2 RPLVLMLGCVLPYGLALKGPKEIGFEGDVTSLACTYERKHKRYKCRGGILVNSC 61
QY 61 ---ILGNQSFLLTKGPSKLNDRADSRSLMDQGNPFLIKNLIKIEDSDTYICEV---D 113
DB 62 GDIVYANQDQEVETGRMSIR---DSPQEL---SMTVIMRDLTLQDSGKVMCGIDRLGHD 114
QY 114 QKEEVQLVVF-----GLTANSDTHLQGGSLTLTLESPSSPVQCRSP-----R 159
DB 115 ESFEVTLVIFPSSYPVWVPLPTTPDS-----RAVSSVSKPSVSIPIVWMAVPLILL 169
QY 160 GKNIOGKTLVSQLELDQSGTGTCTVLONOKK 193
DB 170 SLILAAGLIAFGSHMLRKRKAMLATETQKNEK 203

RESULT 91
Q7ZU00 PRELIMINARY; PRT; 564 AA.
AC Q7ZU00;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Neurotin.
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Body;
RA MEDLINE=22388257; PubMed=12477932;
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buerow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

```

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Guntrane P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huily S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Phelan J., Helton E., Kesteman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywnicki M.I., Skalska U., Smalutz D.E., Schermer A., Schein J.E.,  
RA Jones S.J., Marra M.A.,  
RT "generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Body;  
RA Submitted R.;  
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC050482; AHS0482.1; -  
DR InterPro: IPR003599; Ig\_1ke.  
DR InterPro: IPR007110; Ig\_1like.  
DR InterPro: IPR003598; Ig\_c2.  
DR InterPro: IPR003006; Ig\_MHC.  
DR InterPro: IPR003596; Ig\_v.  
DR Pfam: PF00047; Ig\_1.  
DR SMART: SM00409; IG\_3.  
DR SMART: SM00408; IGC2; 1.  
DR SMART: SM00406; IGv; 1.  
DR PROSITE: PS00835; IG\_LIKE; 3.  
DR PROSITE: PS00290; IG\_MHC; 1.  
SO SEQUENCE 564 AA; 61336 MW; 4E4EE40BE05CA985 CRC64;  
  
Query Match 5.5%; Score 128.5; DB 13; Length 564;  
Best Local Similarity 20.6%; Pred. No. 0.23;  
Matches 92; Conservative 61; Mismatches 137; Indels 157; Gaps 21;  
  
QY 9 HLLVLVLALLPA--TQGN--KVLGKGGDTVELTCTAGSOKS---IQFMKNSNQIKI 61  
DB 2 HSYVLCGATFLAAVAPGSGCLPTVIGYETLIEVPCNNGNNKPDGLIFTKMK--YVKD 58  
QY 62 LGNGSFLTKGSPSKLANDRAUSRSLMDQGNFPIIKNLKIEDSPYICEVEDOKEEYQL 121  
DB 59 DGSFGDLLIKQAOK-----DDPTVSAMDGYKTRV--- 87  
QY 122 VFGLTANSDFHLLQGSGTLTLTLESPGSSPSVQCRSPRKNIGCGKTLVSQLELQDSGT 181  
DB 88 --SIANNS-----SLLIAQGSILTDQRY 107  
QY 182 WTCVVLNOKKVEFKIDIVLAFQKASSIYKKEGEVFSFLAFTVEKLTGSGELMWQ 241  
DB 108 FTCHVAVSSTNLEFSVEVKV--HKKSPAVIKKNVKELENG-----KLTQLGECVVE 157  
QY 242 AERASSSSKSWITPDLKKEVSV-----KRYTQDP-----KIQMGK 276  
DB 158 SANPAADLIW-----MKNNQALVDDQKTIITSDTQKPYGLSTSSRLQYTARKEVYAS 213  
QY 277 KLPPI--HLLPQALPQYASGNITLLEAKTGKHOEVNLVVRATQLOK---NLTCE 329  
DB 214 QFPVAVKHTGPNQV-----STPDTFQIRYPT-----SKVSIQVAVSQSPIREGDVTLLCKQ 264  
QY 330 VMGPTSKMLSLKLENKEAKVSKREKPVWL-----NPEAGMOCCLISDSGOVLLEENIK 385  
DB 265 ADGAPPP--TSFNFNIGKKVYVTDQVYVLTGTVRADSGVAVKCSLLDND--VMESTQI 319  
QY 386 VLPWTSTPVPHPASALPA--PPTGSAL 410  
DB 320 V-----TVSFLDASLPTGKVL 336  
  
RESULT 92  
Q9W213 PRELIMINARY; PRT; 1395 AA.  
ID Q9W213  
AC Q9W213;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE C013521 protein.  
GN ROBO OR CG13521.  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Berkeley;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champs E., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Aoril J.F., Agdayant A., An H.-O., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foele C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegami C.,  
RA Jalali M., Kalish F., Kapten G.H., Ke Z., Kension J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Rainert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spter E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Sytkabas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ananlatides P.G., Brandon R.C., Rogers Y., An H., Baldwin D.,  
RA Banzon J., Beeson K.Y., Busan D.A., Carlson J.W., Center A.,  
RA Champe M., Davenport L.B., Dietz S.M., Dodson K., Dorsett V.,  
RA Doup L.E., Doyle C., Drensek D., Farfan D., Ferreira S., Frise E.,  
RA Galle R.F., Gay N.S., George R.A., Gonzalez M., Houck J.,  
RA Hoskins R.A., Hostin D., Howland T.J., Ibegwam C., Jalali M.,  
RA Kusne D., Li P., Mattei B., Moshrefi A., McIntosh T.C., Moy M.,  
RA Murphy B., Nelson C., Nelson K.A., Nunoo J., Paclet J., Paragas V.,  
RA Puri V., Richards S., Pfeiffer B., Phouanavong S., Pittman G.S.,  
RA Svirskas R., Tector C., Tyler D., Williams S.M., Zaveri J.S.,  
RA Smith H.O., Venter J.C., Rubin G.M.,  
RT "Sequencing of Drosophila melanogaster genome.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RA Hradecky P., Huang Y., Kaminker J.S., Prochownik S.E., Smith C.D.,  
RA Tupy J.L., Bergman C.M., Berman B.P., Carlson J.W., Celniker S.E.,

RA Ciamp M.E., Drysdale R.A., Emmert D., Frise E., de Grey A.D.N.J.,  
 RA Harris N.L., Krommiller B., Marshall B., Milburn G.H., Richter J.,  
 RA Russo S., Searle S.M.J., Smith E., Shu S., Smulnick F.,  
 RA Whitfield E.J., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Mungall C.J., Lewis S.E.,  
 RT "Annotation of Drosophila melanogaster genome."  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AEO03458; AM71113.1; -  
 DR HSP; P56276; 1TLK.  
 DR rlyBase; FBgn0005631; robo.  
 DR GO; GO:0005886; C:plasma membrane; IDA.  
 DR GO; GO:0007411; P:axon guidance; IMP.  
 DR InterPro; IPR003962; FniII subd.  
 DR InterPro; IPR003961; FNI\_III.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR007110; IG\_III-like.  
 DR Pfam; PF00041; fn3; 3.  
 DR Pfam; PF00047; Ig; 5.  
 DR PRINTS; PR00014; FNTYPEIII.  
 DR PROSITE; PS50835; IG\_LIKE; 5.  
 KM Repeat.  
 SQ SEQUENCE 1395 AA; 151759 MW; 25CED7DEB44F13F0 CRC64;

Query Match 5.5%; Score 128.5; DB 5; Length 1395;  
 Best Local Similarity 20.1%; Pred. No. 0.81;  
 Matches 97; Conservative 58; Mismatches 179; Indels 149; Gaps 18;

DB 26 NKVVLGGKDPVELTCTASQKSIQPHMKNSNQIKLNGSFLTGPSKLNRAASRS 85  
 353 NKRV-GING-VQDPCASGNPPSVFWKEGVSTLM-----PPNSHGQHYAAAG--- 402  
 DB LMDQGFPLIKLKIEDSDTYICEVEDQKEVOLVFGLTANSPTN-----LLQGQSLTLT 142  
 403 -----TLQITVRQDEBYVCASFSVDSSTVFLQVSLDERPPIIIIGANGT 455  
 QY 143 LBSPPGSSPSVQCR-----SPRGK-----NIQSKTSLVSQLELDSDGTWT 183  
 456 L--PKGSVATLPCRATGNPSPRIKMFHDGHAVOGNKYSIIQSSSLRVDDLQSLDSGTYT 513  
 DB 184 CTVLQNGKKEFFIDIVLAFQKASSIVYKKEGEQVEFSPFLAFVTEKLTGSELWQME 243  
 514 CTA-----SSEKGETSWAATLTVEK--PGSTSLHRAAD 544  
 DB 244 -----RASSSKSW-----TFPDKNKEVSKY 266  
 QY 545 PSTYAPRPTPKVLANSRTSISLRMAKSOEKGAVGPIIGTYEYSSPDQGTWIAAQR 604  
 DB 267 TQDPKIQMGKLP-----LHLLPQALP-----QYAGSGNLTALDA 303  
 605 VGDQVITISGLTGTSTVFLVRAENTQGISVPSGLSNTIKITIRADPDASANDLSAARTL 664  
 DB 304 KTGKLQEVNLVVMRAQTQKLNLTCEVWGTPSKMLSLKENKEKVSREKPVNLVP 363  
 665 LTGSVELLIDASATNSAVLEWMLHV--SADREYEGELIHYKDAV-----P 711  
 QY 364 EAGWQCLSD--SGQVLLSNIKVLPTWSTVHPRAASALPAPPTGALPDPOTASALPP 422  
 DB 712 SAQYHSTVWDASAESFVGNLKKYKTEFFLTPPFTIEIGQPSNS-----KIALYEDV 766  
 QY 423 PAA 425  
 DB 767 PSA 769  
 RESULT 93  
 08CGB3 PRELIMINARY; PRT; 1413 AA.  
 AC 08CGB3;  
 DT 01-MAR-2003 (TEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)  
 DE Similar to uveal autoantigen with coiled-coil domains and ankyrin repeats.

GN 2700059D02RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; Tissue=salivary gland;  
 RA Strausberg R.;  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC042415; AAH42415.1; -  
 DR MGD; MGI:1919615; 2700059D02RIK.  
 DR InterPro; IPR002110; ANK.  
 DR InterPro; IPR000727; T\_SNARE.  
 DR Pfam; PF00023; ank; 5.  
 DR SMART; SM00248; ANK; 6.  
 DR PROSITE; PS50088; ANK\_REPEAT; 5.  
 DR PROSITE; PS50297; ANK\_REPEAT\_REGION; 1.  
 DR PROSITE; PS50192; T\_SNARE; 1.  
 SQ SEQUENCE 1413 AA; 160917 MW; 2DE27904AF684120 CRC64;

Query Match 5.5%; Score 128.5; DB 11; Length 1413;  
 Best Local Similarity 20.5%; Pred. No. 0.83;  
 Matches 76; Conservative 72; Mismatches 144; Indels 79; Gaps 12;

QY 4 GVP-----FRHLIVQLALPAAQGNKVLGKGGDTVELTCTASQKSIQPHMKNSNQI 59  
 DB 419 GTPHMHSSMKRPLELSPSOTSYSENEILKKELETRTTYDSAKQDRKQ----- 471  
 QY 60 KILNGSFLTGPSKLNDRADRSRLMDQGNFPLIKLKIEDSDTYICEVEDQKEVQ 119  
 DB 472 -----NELAHKVAECKAL-----ALECEKVK--EDSDQIKQEDALDKVQ 510  
 QY 120 LLVF-----GLTANSPTNHLQGSGLTLTLESPSSSVQCRSGRNIG-----GKTLVSQ 173  
 DB 511 KMYSESEGVKOMQHPALKE--HLTNEPATSHRIIELRQDLKQKTEGASAEVQK 569  
 QY 174 LELQ-----DSGTWTCTVLQNGKKEFFIDIVLAFQKASSIVYKKEGEQVE 220  
 DB 570 LASQIKQSEMLVGEKRDG-----RLVEENRLQKECGCEVELRGRGVVLELQK 625  
 QY 221 FSPFLAFVTEKLTGSGELWQERASSKSWITFDLKNKEVSKVATQDPKIQMGKLP 280  
 DB 626 LQAKLALSV-----PTEKFSMSKSLSDINERKVLAEVGRDYESAQGEIRQL 674  
 QY 281 HLLTQALPQY---AGSGNLTALAEKTKLQEVNLVVMRAQTQKLNLTCEVWGTPSK 337  
 DB 675 KQDLESVPAQIHREPHEQLRSRLQKSGELKQVSELTQKQTLQKQV-----BK 725  
 QY 338 LMLSLKENKE 348  
 DB 726 LHADNKLNLNQ 736  
 RESULT 94  
 008653  
 ID 008653 PRELIMINARY; PRT; 2629 AA.  
 AC 008653;  
 DT 01-JUL-1997 (TEMBLrel. 04, Created)  
 DT 01-JUL-1997 (TEMBLrel. 04, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE Telomerase protein component 1.  
 GN TLPI.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 ON NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Fischer 344;  
 RX MEDLINE=97236507; PubMed=9118230;  
 RA Nakayama J., Saito M., Nakamura H., Matsuura A., Ishikawa F.,  
 RT "TLPI: a gene encoding a protein component of mammalian telomerase is

RT a novel member of WD repeats family."  
 RL Cell 88:875-884(1997).  
 DR EMBL: U89282; AAB51690.1; -.  
 DR PIR: T32735; T32735.  
 DR InterPro: IPR007111; NACHT\_NTPase.  
 DR InterPro: IPR008850; TEPI\_N.  
 DR InterPro: IPR008858; TROVE.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF05729; NACHT; 1.  
 DR Pfam: PF05386; TEPI\_N; 4.  
 DR Pfam: PF05731; TROVE; 1.  
 DR Pfam: PF00400; WD40; 15.  
 DR Prodom: PD000018; WD40; 1.  
 DR SMART: SM00320; WD40; 15.  
 DR PROSITE: PSS0837; NACHT; 1.  
 DR PROSITE: PSS0678; WD\_REPEATS\_1; 1.  
 DR PROSITE: PSS0082; WD\_REPEATS\_2; 7.  
 DR PROSITE: PSS0294; WD\_REPEATS\_REGION; 2.  
 KM Repeat: WD repeat.  
 SQ SEQUENCE 2629 AA; 291705 MW; B90C3F72E5D4B6E CRC64;

Query Match 5.4%; Score 127; DB 11; Length 2629;  
 Best Local Similarity 20.3%; Pred. No. 2.6;  
 Matches 88; Conservative 62; Mismatches 133; Indels 150; Gaps 20;

QY 9 HLLVQLAL-----PAATQGNKVVLGGKGDVLE-----TCTASQKSI- 49  
 DB 2235 HPLVLCIHITLQGHSGVPTAAASGALLTSDNSSVRLMOJPKKADTCKPRSSAVIT 2294  
 QY 50 QFHWKNSQIKILNCGSFLL-----KGPSKLAND-----RADSRSLL 86  
 DB 2295 AVAMAPPGSLVSNENAGELTLMQKAVATAPRAGVSLDINCSANAFVLANEVSE 2254  
 QY 87 WD-----QGFLLIKKLIKIEDSDTYICEVEDQKEVQLVGLTANSSTHLLQGS 138  
 DB 2355 MOVELRKGSTCTNRLYLKVLQEDLG-----VLTMALAPD-----GQS 2394  
 QY 139 LTL-----TLESPPGSSPVQCGSPRGKN-----IQGKTLVSQLEIQDS 179  
 DB 2395 LILMKEDVELLMKPGSTPSSICRRYAVHSSILCTSDYGLFYIQGNSGSLSTLEDES 2454  
 QY 180 GTWTCVTLQOKKVEFK-----IDIVLAFQKASSIVYKKEGEQVEFSPPLAFV 229  
 DB 2455 GKF-----EKTLDFYLNINPNPGSPVSTIQAEPEGSSSL-----CATSGMLMNL 2500  
 QY 230 EKLTSGE-----LWQAERASSK-----SWITFDLKNKEVSVKRVTD 270  
 DB 2501 SECTPEGEWVVDINWQKSRNPKSRTPGTDSPGLFCMDSWV-----EPTHLKARQCK 2553  
 QY 271 KLGKGLPLHLTLPQALPQYAGSNLTALFATGKLGHOENVLVNRAATOLQKNTLC-E 329  
 DB 2554 KIHLSGYTALHV-----LP-----GLLVASSEDVATLMERPQQLLGLRCBPVSCLE 2603  
 QY 330 VMGPTSPKLMSTL 342  
 DB 2604 PMMERPSPLQALV 2616

RESULT 95  
 Q9ESB5 PRELIMINARY; PRT; 399 AA.  
 AC Q9ESB5;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE NK cell receptor 284.  
 GN NMRK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=BALB/c;  
 RX MEDLINE=20395298; PubMed=10941850;  
 RA Kumaresan P.R., Huynh V.T., Mathew P.A.;  
 RT "Polymorphism in the 2B4 gene of inbred mouse strains."  
 RL Immunogenetics 51:758-761(2000).  
 DR EMBL: AF184208; AAG12323.1; -.  
 DR MGD; MGI:109294; Nmrk.  
 DR GO; GO:0004872; F:receptor activity; IEA.  
 DR InterPro: IPR003599; IG.  
 DR InterPro: IPR007110; IG-like.  
 DR SMART: SM00409; IG; 1.  
 KM Receptor.  
 SQ SEQUENCE 399 AA; 44991 MW; 632E61217FE1D065 CRC64;

Query Match 5.4%; Score 126.5; DB 11; Length 399;  
 Best Local Similarity 21.3%; Pred. No. 0.2;  
 Matches 84; Conservative 61; Mismatches 151; Indels 99; Gaps 16;

QY 18 LLEPATQGN-----KVLGKKGDVLELTCTASQKSIQFHWKNSQIKILNCGSFLLTK 71  
 DB 12 LILRAHQGDGDSPREEVGVSGKPYRLRPSNIQTQDVAVQMKK-----KVQSHNM--TE 65  
 QY 72 GPGKLANDRADSRSRL-----WDQGNFPLIKLIKIEDSDTYICEVEDQKEV----- 118  
 DB 66 ILISCNNA-SRCSIVSSDIYGFYGDPALSIKSAQLDSCGHYLLLETYRMGTCTKNF 124  
 QY 119 QLLVFLTANSSTHLLQGSLLTLESPPGSSPVQCGSPRGKNIO-----GKTLVS-- 172  
 DB 125 QILILD---HVEFPHLKAQNMAMTNGT---COLFLSCVSPKQDNVYALYRGSTLMIISNQ 178  
 QY 173 -----OLELDGSGTWTCVTLQ-----NOKKVEFKDIVLAF 204  
 DB 179 RNSITHENQTDASLHTYTCNNSNRASMANHTLNFHGCOSVHLNRLPFGYIYIIVLT 238  
 QY 205 QKASSI-----VYKKEGEQVEFSPPLAFVYEKLTGSGELMWQAE--RASSSKSWITFDLK 257  
 DB 239 LFLGALICFCVWTKKRRKQLOFSPKEPLTYEVYKDSRASRDQCGSRAGSPSDVQEDGR 298  
 QY 258 NKEVSVKRVTDQPKLQMGKPLHLTLPQALPQYAGSNLTALFATGKLGHOENVLVN 317  
 DB 299 GQRELDNRVSE-----VLBQLPQNSPGG-----RGTMYSMIOCKPE 335  
 QY 318 RATOLQKNTLCEVWGPSPPLMSTLKLENKEAVS 352  
 DB 336 DSTSQEK---CTYISVYQPSRKSGSKKRNQNSLS 367

RESULT 96  
 Q26475 PRELIMINARY; PRT; 484 AA.  
 AC Q26475;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE REGA-1 protein precursor.  
 GN REGA-1.  
 OS Schistocerca americana (American grasshopper).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Orthopteroidea; Orthoptera; Caellifera; Acridomorpha;  
 OC Acridoidea; Acrididae; Cyrtacanthacridinae; Schistocerca.  
 OX NCBI\_TaxId=7009;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96189346; PubMed=8625808;  
 RA Seaver E.C., Carpenter E.M., Bastiani M.U.;  
 RT "REGA-1 is a GPI-linked member of the immunoglobulin superfamily  
 RT present on restricted regions of sheath cell processes in  
 RT grasshopper."  
 RL Development 122:567-578(1996).  
 DR EMBL: X93601; CAA63800.1; -.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003598; IG\_c2.





OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ruddy D.A., Krommal G.S., Lee V.K., Mintier G.A., Quintana L.,  
 RA Domingo R. Jr., Meyer N.C., Basava A., McClelland E., Fullan A.,  
 RA Mapa F.A., Moore T., Thomas W., Loeb D.B., Harmon C., Tsuchishashi Z.,  
 RA Wolfe R.K., Schatzman R.C., Feder J.N.;  
 RT "A 1.1 megabase transcript map of the human hereditary hemochromatosis  
 locus";  
 RL Submitted (FE8-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U90552; AAB53430.1; -;  
 DR Genem; HGNC:1138; BTN3A1.  
 DR GO; GO:0006629; P:lipid metabolism; TAS.  
 DR InterPro; IPR001870; B302.  
 DR InterPro; IPR003599; IG.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR006574; PRY.  
 DR InterPro; IPR003877; SPRY\_receptor.  
 DR Pfam; PF00622; SPRY, 1.  
 DR SMART; SM00409; IG, 1.  
 DR SMART; SM00589; PRY, 1.  
 DR SMART; SM00449; SPRY, 1.  
 DR PROSITE; PS50835; IG\_LIKE, 2.  
 DR PROSITE; PS50835; IG\_LIKE, 2.  
 SQ SEQUENCE 513 AA; 57762 MW; CD334D727CD1F63 CRC64;

Query Match 5.4%; Score 125.5; DB 4; Length 513;  
 Best Local Similarity 21.3%; Pred. No. 0.34;  
 Matches 102; Conservative 72; Mismatches 191; Indels 113; Gaps 25;

QY 7 FRHLVLVQLALPAATQGNKV-----VLGKGGDTVELTC--TASQKSIQFHWKNSQ 58  
 DB 14 FRVCLLLQLQ-LMPHSQFSLGPSGPIILMVGEDADLPCHLPPTMSAETMELKVVSSSL 72  
 QY 59 IKLL-----GNQSFLLTKGSKLNDPDRSRSLMDQGNFPLIKLKIHSPTIYCEVD 113  
 DB 73 RQVNVVYADGEVEDRQSAFYR--GRTSILRDGITGAKALRIHNVTASDSGKYLQYFQD 130  
 QY 114 -----OKEEVQLVFGLTANSDFHL-----LQGSLLTLTLESPPG--SSPSVQCRSPRGKNIQ 164  
 DB 131 GDYEKALVLEKLAAL--GSDLDVQVGYDGGSIHLECRSTGWYPPQOLQMSNNKENIP 188  
 QY 165 -----GKTLVSQLELQDSGTW--TCTVLQNKQKVEPKIDIVLAFQKASSI-- 210  
 DB 189 TVEAPVADGVGLYVAASVIMRSGSGEVSTIRSS-----LLGLEKTASIS 237  
 QY 211 --VYKKEGEG--VEFSPLAFVTEKLTGSGELMWQAEKASSKSWITFDLKNKEVSKV 266  
 DB 238 ADFFPSAQKWIALATLFLVLLLLGAGYFLMQOEEKTQ---FRKKKEQELREH 293  
 QY 267 T-QDPKLMQKGLPLHLTLPLQALPOYAGSGN-----LTIALAKTGK-- 307  
 DB 294 ANSTMQEOSTRKYLLEELMRSIQYASRGERHSAYNEMKALPKRADVILDKXTANPIL 353  
 QY 308 LHQEVLLVVRATQLOK-----NLTCVWGPTSPKMLSLKL-ENKEAKVSK 353  
 DB 354 LVSEDDQSRVQRAKEPQLPDPNPERFMHYCVLGCESEF--ISGRHYWEVEVGDRKEWHIGV 411  
 QY 354 REKPV---WV-LNPELGMWQCLSDSGQVLLSNIKVLPWTSTPVNHRAS-ALPAP 405  
 DB 412 CSKNVQRKGVKMTPENGFTWGLTDGK-----YRTLTPEPTNLKLPKP 457

RESULT 100

O921P2 PRELIMINARY; PRT; 605 AA.  
 AC O921P2;  
 DT 01-DEC-2001 (TEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN NCAM1.  
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Streusberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC011310; AAH11310.1; -;  
 DR MGD; MGI:97281; Ncam1.  
 DR InterPro; IPR003961; FN\_III.  
 DR InterPro; IPR008957; FN\_III-like.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003598; IG\_c2.  
 DR Pfam; PF00041; fn3, 1.  
 DR Pfam; PF00047; IG, 5.  
 DR SMART; SM00060; FN3, 1.  
 DR SMART; SM00408; IGC2, 5.  
 DR PROSITE; PS50835; IG\_LIKE, 5.  
 KW Hypothetical protein; Immunoglobulin domain.  
 SQ SEQUENCE 605 AA; 67353 MW; DED98C64D245A867 CRC64;

Query Match 5.4%; Score 125; DB 11; Length 605;  
 Best Local Similarity 19.9%; Pred. No. 0.47;  
 Matches 61; Conservative 48; Mismatches 139; Indels 58; Gaps 12;

QY 3 RG-VFRRHLVLVQLALPAATQGNKVVLGKGGDTVELTCTASQKSIQFHW-KNSNQIK 60  
 DB 196 REINFPKQIIVNVNPPYVQARQSIYNATANQGSVTLVCDADGPREPTMSWTKGEPLE 255  
 QY 61 ILGNQSFLLTKGSKLNDPDRSRSLMDQGNFPLIKLKIHSPTIYCEV---DQKE 116  
 DB 256 -----NEEDDEKHI FSDSSELTIRVNDKNDEAEVYCIENKAGEODA 299  
 QY 117 EYQLVFG--LTANSPTHLQGSLLTLTLESPPGSSPSVQCR-----SPRGKNIQ 166  
 DB 300 STLKLVFAKPKITTYENQTALEEQVTLTCEASGDPISITWRTSTRNISSEKTLDOH 359  
 QY 167 -----KTLVSQLELQDSGTWCTVLQ--NOKVVEPKIDIVLAFQKASSIVYKKE 215  
 DB 360 MVVRSHARVSLTLKSIQYTDAGEYICTASNTIGDSQSMYLEVGYAPRLQGPVAVYWE 419  
 QY 216 GEQV---SFPLAFVTEKLTGSGELMWQAEKASSKSWITFDLKN-KEVSKRYTOD 269  
 DB 420 GNQVITCEVFAYPSA-TIS-----WFRDQQLPSSNYSNIXITYTPPSASYLEVTPD 470  
 QY 270 PKLQMG 275  
 DB 471 SENDFG 476

Search completed: August 3, 2004, 13:13:35  
 Job time : 38.1057 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:50:09 ; Search time 46.0872 Seconds  
(without alignments)  
2777.216 Million cell updates/sec

Title: SEQ5  
Perfect score: 2325  
Sequence: 1 MNRGVPRHLLVLQLALP.....VISFLGLGLGVACVLARTR 453

Scoring table: BLOSUM62

Searched: 1586107 beqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 125 summaries
```

Database : A\_GeneSeq\_29Jan04:

1: genebeqp1980s: \*  
2: genebeqp1990s: \*  
3: genebeqp2000s: \*  
4: genebeqp2001s: \*  
5: genebeqp2002s: \*  
6: genebeqp2003as: \*  
7: genebeqp2003bs: \*  
8: genebeqp2004s: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2049	88.1	474	3	AAV59170	AAV59170 CD4-Ig fu
2	2047.5	88.1	436	3	AAV51080	AAV51080 Human fu
3	2046	88.0	458	3	AA807769	AA807769 DNA encoc
4	2044.5	87.9	729	3	AA813507	AA813507 CD4-IgG1
5	2042	87.8	729	1	AAV93008	AAV93008 Genetic C
6	2042	87.8	729	3	AAV51078	AAV51078 Human fu
7	2042	87.8	729	3	AAV59168	AAV59168 CD4-Ig fu
8	2040.5	87.8	416	3	AA81509	AA81509 CD4-IgM f
9	2040	87.7	462	2	AA821277	AA821277 CD4-εeta f
10	2040	87.7	462	2	AA876677	AA876677 T-cell re
11	2040	87.7	462	2	AA8989457	AA8989457 CD4:Gamma
12	2040	87.7	462	2	AA802214	AA802214 CD4:Fc re
13	2040	87.7	462	2	AA898142	AA898142 ChimERIC
14	2038.5	87.7	575	2	AA821276	AA821276 CD4:zeta
15	2038.5	87.7	575	2	AA876676	AA876676 T-cell re
16	2038.5	87.7	575	2	AA898456	AA898456 CD4:zeta
17	2038.5	87.7	575	2	AA802213	AA802213 CD4:T-cel
18	2038.5	87.7	575	2	AA891990	AA891990 ChimERIC
19	2038	87.7	458	1	AA811990	AA811990 Clone pr
20	2038	87.7	458	1	AAAP91369	AAAP91369 T4 prote
21	2038	87.7	458	2	AAV38826	AAV38826 Soluble f
22	2038	87.7	481	3	AA819510	AA819510 CD4-IgM f
23	2038	87.7	481	3	AAV59171	AAV59171 CD4-Ig fu
24	2037	87.6	398	2	AA898450	AA898450 CD4-DL
25	2037	87.6	532	2	AA821278	AA821278 CD4:Gamma

```

99 1872 80.5 370 2 AAR12967 Asp(155),
100 1871 80.5 435 1 AAP91289 Sequence
101 1870 80.4 370 2 AAR12955 Gly(48),
102 1869 80.4 370 2 AAR12963 His(132),
103 1868 80.4 370 2 AAR12960 Lys(99),
104 1867 80.3 433 2 AAR08335 Ser(121),
105 1866 80.3 370 2 AAR12961 Ser(121),
106 1862 80.1 458 2 AAR10987 Rheus mo
107 1861 80.0 458 2 AAR20148 Cynomolgu
108 1860 80.0 370 2 AAR12953 Thr(127),
109 1859 80.0 400 2 AAR20149 Sol. rhe
110 1858 58.8 315 2 AAR07606 Plaemid T
111 1363 58.6 295 2 AAR07605 C-termina
112 1262 54.3 273 2 AAR21078
113 1113.5 47.9 454 2 AAR33181 L374 muta
114 1080 46.5 507 2 AAR54201 Feline T
115 1059 45.5 474 2 AAB00158 BCD4-SCFV
116 1050.5 45.2 310 2 AAR26784 CD4-IgG2
117 1050.5 45.2 310 2 AAR46680 CD4-kappa
118 1050.5 45.2 310 3 AAR85081 CD4-kappa
119 1050.5 45.2 310 4 AAB67324 CD4-kappa
120 1050.5 45.2 310 4 AAB80885 Human CD4
121 1050.5 45.2 310 6 AAB71124 CD4-kappa
122 1048 45.1 534 2 AAR26531 Sequence
123 1046 45.0 530 2 AAR26783 CD4-IgG2
124 1046 45.0 530 3 AAR85080 CD4-IgG2
125 1046 45.0 530 4 AAB67323 CD4-IgG2

```

## ALIGNMENTS

RESULT 1  
ID AAY59170 standard; protein; 474 AA.

AC AAY59170;

DT 14-MAR-2000 (first entry)

XX CD4-Ig fusion protein CD4Mmu.

XX HIV; extracellular; CD4; gp120; immunoglobulin; Ig; fusion protein;

KM secreted protein; HIV infection; medicament.

OS Synthetic.

OS Homo sapiens.

PN CA1340741-C.

PD 14-SEP-1999.

PF 20-JAN-1989; 89CA-00588749.

PR 20-JAN-1989; 89CA-00588749.

PA (GEHO ) GEN HOSPITAL CORP.

PI Seed B;

DR WPI; 2000-063015/06.

DR N-PSDB; AA248203.

PT New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in  
the treatment of HIV or simian immunodeficiency virus infections.

XX Example 1; Page 47-53; 89pp; English.

XX The invention provides a fusion gene encoding a fusion protein that  
comprises an extracellular CD4 DNA sequence or its fragment which binds  
to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA  
sequence of an Ig heavy or light chain, where the DNA sequence encoding  
the variable region has been replaced with the DNA sequence which encodes

CC extracellular CD4 or its gp120 binding fragment. The fusion protein is  
CC capable of being secreted. The fusion proteins are useful for treating  
CC HIV or HIV infections in animals, preferably humans. They are also useful  
CC for producing medicaments which can be used for treating HIV or HIV  
CC infections in humans. The present sequence represents the fusion protein  
CC CD4Mmu where the CD4 is linked to human IgG1 at the Mat2 site upstream of  
CC the CH1 region  
XX

SO Sequence 474 AA;

Query Match 88.1%; Score 2049; DB 3; Length 474;

Best Local Similarity 91.6%; Pred. No. 1,le-131;

Matches 402; Conservative 6; Mismatches 21; Indels 10; Gaps 1;

```

QY 1 MNRGVPRHLVLYQLALPAATQGNKVVGGKGDVLELTCTASQKSIQPHMNSQIK 60
DB 1 MNRGVPRHLVLYQLALPAATQGNKVVGGKGDVLELTCTASQKSIQPHMNSQIK 60
QY 61 ILGNQGSFLTKGPKLNDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGPKLNDRADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSSTHLLQGSILTLTLESPGSSPSVQCRSPRKNIOGKTLISVQLDSDG 180
DB 121 LVFGLTANSSTHLLQGSILTLTLESPGSSPSVQCRSPRKNIOGKTLISVQLDSDG 180
QY 181 TWTCITVLONQKQVEFKIDIVLAFQKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMW 240
DB 181 TWTCITVLONQKQVEFKIDIVLAFQKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMW 240
QY 241 QABRASSSKSWITFDLKNKEVSVKVYTOPDKLQWKKLPHLTLPOALPOYAGSGNLTLLA 300
DB 241 QABRASSSKSWITFDLKNKEVSVKVYTOPDKLQWKKLPHLTLPOALPOYAGSGNLTLLA 300
QY 301 LEATGKGLHOBVLVVMRATQLOKLTCEVWGPSPPLMLSLKLENKAVSKREKEVWV 360
DB 301 LEATGKGLHOBVLVVMRATQLOKLTCEVWGPSPPLMLSLKLENKAVSKREKEVWV 360
QY 361 LNPAGMOCILSDSGVLESNIKVLPWTSTPV-----HPRASALPAPPTGSAL 410
DB 361 LNPAGMOCILSDSGVLESNIKVLPWTSTPVHADPELPLTAHPGQTLHSSDTCR 420
QY 411 PDQTSALPDPAPASALP 429
DB 421 PKTPKAKLSTPSAKRPAP 439

```

RESULT 2

ID AAY51080 standard; protein; 436 AA.

AC AAY51080;

DT 23-MAR-2000 (first entry)

XX Human fusion protein CD4Mg.

XX Fusion protein; human; CD4; IgM; immunoglobulin; gp120;

KM anti-human immunodeficiency virus; CD4Mg.

OS Homo sapiens.

OS Synthetic.

PN US6004781-A.

PD 21-DEC-1999.

PF 04-FEB-1994; 94US-00191708.

PR 22-JAN-1988; 88US-00147151.

PR 23-JAN-1989; 89US-00299596.

PR 09-JUN-1992; 92US-00896781.

PR 12-APR-1993; 93US-00057952.

```

XX (GEHO ) GEN HOSPITAL CORP.
PA
XX
PI Seed B;
PI
XX WPI; 2000-085792/07.
DR N-PSDB; AA244063.
XX
PT Fusion protein useful for the treatment of human immunodeficiency virus.
XX
PS Example 1; Col 41-50; 39pp; English.
XX
CC This invention describes a novel nucleic acid (I) encoding a fusion
CC protein comprising a DNA sequence encoding amino acids 1-173 of CD4 (II)
CC and a DNA sequence encoding a human immunoglobulin (Ig) heavy or light
CC chain (III). The products of the invention have anti-human
CC immunodeficiency virus (HIV) activity and are capable of binding to
CC gp120. The fusion protein is useful for treating human immunodeficiency
CC virus (HIV) or simian immunodeficiency virus (SIV). This sequence
CC represents the fusion protein CD4g which is constructed from CD4 linked
CC to human IgM upstream of the CH1 region
XX
SQ Sequence 436 AA;

Query Match      88.1%; Score 2047.5; DB 3; Length 436;
Best Local Similarity 93.0%; Pred. No. 1.3e-131;
Matches 401; Conservative 3; Mismatches 12; Indels 15; Gaps 1;

QY 1 MNRGVFPHLLLVQLALLPAATQGNKVVYLGKKGDTVELTCTASQKKSIOFHMKNSNOIK 60
DB 1 MNRGVFPHLLLVQLALLPAATQGNKVVYLGKKGDTVELTCTASQKKSIOFHMKNSNOIK 60
QY 61 ILGNQGSFLTKGPKSKLNDRADSRSLMDQGNFPLIKNLIKIEDSDTYICEVEDQKEEYOL 120
DB 61 ILGNQGSFLTKGPKSKLNDRADSRSLMDQGNFPLIKNLIKIEDSDTYICEVEDQKEEYOL 120
QY 121 LVFGLTANSPTHLLQGSQSLTTLTLESPGSSPVQCRSPRGKNIOGKTLVSQLELDQSG 180
DB 121 LVFGLTANSPTHLLQGSQSLTTLTLESPGSSPVQCRSPRGKNIOGKTLVSQLELDQSG 180
QY 121 LVFGLTANSPTHLLQGSQSLTTLTLESPGSSPVQCRSPRGKNIOGKTLVSQLELDQSG 180
DB 121 LVFGLTANSPTHLLQGSQSLTTLTLESPGSSPVQCRSPRGKNIOGKTLVSQLELDQSG 180
QY 181 TWTCTVLQNOQKKEFKIDIVVLAFOKASSIYVKKGEQVEFSPLAFTVEKLTGSGELMW 240
DB 181 TWTCTVLQNOQKKEFKIDIVVLAFOKASSIYVKKGEQVEFSPLAFTVEKLTGSGELMW 240
QY 241 QABERASSSSWITFDLAKKEVSKVKTQOPKLOMGKKLPLHLTPLPALQVYAGSGMLTIA 300
DB 241 QABERASSSSWITFDLAKKEVSKVKTQOPKLOMGKKLPLHLTPLPALQVYAGSGMLTIA 300
QY 301 LEAKTGKLEHOEVNLVVMRATOLQKNLTCEVMGPTSPKMLSLKLENKEAKVSKREKPVWV 360
DB 301 LEAKTGKLEHOEVNLVVMRATOLQKNLTCEVMGPTSPKMLSLKLENKEAKVSKREKPVWV 360
QY 361 LNPEAGMOCCLSDSGQVLLSNIKVLPTWSTPVHPRASALPAPPTGSAIPDOTASALP 420
DB 361 LNPEAGMOCCLSDSGQVLLSNIKVLPTWSTPVH-----ADPECCXPKP 405
QY 421 DEPPASALPAA 431
DB 406 TPKAKLSTPSA 416

RESULT 3
AAB07769
ID AAB07769 standard; protein; 458 AA.
XX
AC AAB07769;
XX
DT 07-NOV-2000 (first entry)
XX
DE DNA encoding a human T4 glycoprotein.
XX
KW Human; T4 glycoprotein; human immunodeficiency virus; HIV;
envelope glycoprotein; AIDS; virus binding.

```

```

XX OS Homo sapiens.
XX
XX FH Key
XX FT Peptide
XX FT Location/Qualifiers
XX FT 1..23
XX FT /note="leader sequence"
XX FT Modified-site
XX FT 296..298
XX FT /note="N-linked glycosylation site"
XX FT Modified-site
XX FT 325..327
XX FT /note="N-linked glycosylation site"
XX FT Domain
XX FT 398..420
XX FT /note="transmembrane domain"
XX FT Domain
XX FT 421..458
XX FT /note="cytoplasmic domain"
XX
XX PN US6093539-A.
XX
XX PD 25-JUL-2000.
XX
XX 06-JUN-1995; 95US-00466368.
XX
XX 21-AUG-1986; 86US-00898587.
XX 11-JUN-1991; 91US-00713564.
XX 06-JUL-1992; 92US-00909021.
XX 12-DEC-1994; 94US-00354452.
XX
XX (UYCO ) UNITV COLUMBIA NEW YORK.
XX
XX PA Maddon PJ, Chess L, Axel R, Weiss R, McDougal JS, Littman DR;
XX PI WPI; 2000-505203/45.
XX DR N-PSDB; AA59352.
XX
XX DR New isolated nucleic acid encoding a human T cell surface protein and the
XX PT soluble surface T4 glycoprotein that it encodes, useful as prophylaxis
XX PT for treating a subject infected with human acquired immune deficiency
XX PT syndrome virus.
XX
XX PS Disclosure; Fig 6A-B; 69pp; English.
XX
XX CC The present sequence represents a human T4 glycoprotein. An aqueous-
XX CC soluble polypeptide comprising a portion of a human T4 glycoprotein
XX CC specifically forms a complex with a human immunodeficiency virus (HIV)
XX CC envelope glycoprotein. The DNA is useful for producing the soluble
XX CC surface T4 glycoprotein. The soluble surface T4 glycoprotein is useful as
XX CC a therapeutic agent, i.e. as prophylaxis for treating a subject infected
XX CC with an HIV virus. Thus, the soluble T4 glycoprotein is useful for
XX CC treating human AIDS. The soluble T4 glycoprotein is also useful in
XX CC diagnostic or screening assays, e.g. for screening inhibitors of virus
XX CC binding, or for detecting and quantitating T4, T4+ cells and antibodies
XX CC to T4, which are of diagnostic value for AIDS
XX
XX SQ Sequence 458 AA;

Query Match      88.0%; Score 2046; DB 3; Length 458;
Best Local Similarity 89.6%; Pred. No. 1.7e-131;
Matches 407; Conservative 2; Mismatches 15; Indels 30; Gaps 2;

QY 1 MNRGVFPHLLLVQLALLPAATQGNKVVYLGKKGDTVELTCTASQKKSIOFHMKNSNOIK 60
DB 1 MNRGVFPHLLLVQLALLPAATQGNKVVYLGKKGDTVELTCTASQKKSIOFHMKNSNOIK 60
QY 61 ILGNQGSFLTKGPKSKLNDRADSRSLMDQGNFPLIKNLIKIEDSDTYICEVEDQKEEYOL 120
DB 61 ILGNQGSFLTKGPKSKLNDRADSRSLMDQGNFPLIKNLIKIEDSDTYICEVEDQKEEYOL 120
QY 121 LVFGLTANSPTHLLQGSQSLTTLTLESPGSSPVQCRSPRGKNIOGKTLVSQLELDQSG 180
DB 121 LVFGLTANSPTHLLQGSQSLTTLTLESPGSSPVQCRSPRGKNIOGKTLVSQLELDQSG 180
QY 121 LVFGLTANSPTHLLQGSQSLTTLTLESPGSSPVQCRSPRGKNIOGKTLVSQLELDQSG 180
DB 121 LVFGLTANSPTHLLQGSQSLTTLTLESPGSSPVQCRSPRGKNIOGKTLVSQLELDQSG 180
QY 181 TWTCTVLQNOQKKEFKIDIVVLAFOKASSIYVKKGEQVEFSPLAFTVEKLTGSGELMW 240
DB 181 TWTCTVLQNOQKKEFKIDIVVLAFOKASSIYVKKGEQVEFSPLAFTVEKLTGSGELMW 240

```

```

QY 241 QAERASSKSMITFDLKNKEVSVKRVTOPDKLQMGKLLPLHLTLPOALPOYAGSGNLTLA 300
    |||||
DB 241 QAERASSKSMITFDLKNKEVSVKRVTOPDKLQMGKLLPLHLTLPOALPOYAGSGNLTLA 300
QY 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKLMSTLKENKEAVSKKEKPVWV 360
    |||||
DB 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKLMSTLKENKEAVSKKEKPVWV 360
QY 361 LNPEAGMOCCLSDSGVLLSESNIKVLPWTSTPVHAPRALPAPTGSALDPQTASALP 420
    |||||
DB 361 LNPEAGMOCCLSDSGVLLSESNIKVLPWTSTPVHAPRALPAPTGSALDPQTASALP 420
QY 421 DPPASALPALAVISFLIGLGLGY-ACVLARTR 453
    |||||
DB 397 -----MALIVLGVAGVAGLLFTIGLGIFFCVRCRHR 425

RESULT 4
AAB19507
ID AAB19507 standard; protein; 729 AA.
XX
AC AAB19507;
XX
DT 09-JAN-2001 (first entry)
XX
DE CD4-IgG1 fusion protein CH4Hgamma1.
XX
KM CD4; IgG1; human; CD4Hgamma1; fusion protein; immunoglobulin; HIV; SIV;
KW gp120; therapy; diagnosis.
XX
OS Homo sapiens.
XX
FH Key 1.395 Location/Qualifiers
FT Protein /note="CD4 extracellular region"
FT Protein 400..729
FT /note="IgG1 heavy chain"
XX
PN US6117656-A.
XX
PD 12-SEP-2000.
XX
PF 07-JUN-1995; 95US-00479353.
XX
PR 22-JAN-1988; 88US-00147351.
PR 23-JAN-1989; 89US-00295356.
PR 09-JUN-1992; 92US-00896781.
PR 12-APR-1993; 93US-00057952.
PR 04-FEB-1994; 94US-00191708.
XX
PA (GENO ) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
DR WPI; 2000-586558/55.
DR N-PSDB; AAA50660.
XX
PT CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or
    SIV.
XX
PS Example 1; Col 13-30; 39pp; English.
XX
CC The present sequence is that of fusion protein CD4Hgamma1 comprising the
    extracellular portion of CD4, which binds to HIV gp120, linked at its C-
    terminus to the human IgG1 heavy chain. To obtain the fusion protein, DNA
    encoding CD4 was linked to IgG1 DNA at the Hind3 site upstream of the CH1
    region (see AAA50660). Fusion protein CD4Hgamma1 and a nucleic acid
    encoding it are claimed. Also claimed are a vector comprising the nucleic
    acid, and a method of producing the fusion protein in secreted form using
    a transformed host cell. The fusion protein may further comprise a
    therapeutic agent, radiolabel or NMR imaging agent. The fusion protein
    can be administered to an animal (including humans) for treatment of HIV

```

```

CC or SIV infection, and can also be used in assays for HIV or SIV, imaging
CC and tissue stains. IgG1 fusion proteins such as CD4Hgamma1 provide both
CC complement-mediated and cell-mediated immunity
XX
SQ Sequence 729 AA;
Query Match 87.9%; Score 2044.5; DB 3; Length 729;
Best Local Similarity 91.5%; Pred. No. 3.8e-131;
Matches 407; Conservative 3; Mismatches 22; Indels 13; Gaps 3;

QY 1 MNRGVPFRHLVLLQALALPRAATQGNKVLGKGDVLELTCTASQKKSIOFHMKNNOIK 60
    |||||
DB 1 MNRGVPFRHLVLLQALALPRAATQGNKVLGKGDVLELTCTASQKKSIOFHMKNNOIK 60
QY 61 ILNGQSFLLTKGSPKLNDRADSRSLMDQNPFLITNKLIEBDDTYICEVEDQKEEVOL 120
    |||||
DB 61 ILNGQSFLLTKGSPKLNDRADSRSLMDQNPFLITNKLIEBDDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGSLLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELQDSG 180
    |||||
DB 121 LVFGLTANSDTHLLQGSLLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELQDSG 180
QY 181 TWTCVTLQNGKVEFKIDIVLAPQKASSIVYKKEGQVEFSPPLAFTVEKLTSGGLMW 240
    |||||
DB 181 TWTCVTLQNGKVEFKIDIVLAPQKASSIVYKKEGQVEFSPPLAFTVEKLTSGGLMW 240
QY 241 QAERASSKSMITFDLKNKEVSVKRVTOPDKLQMGKLLPLHLTLPOALPOYAGSGNLTLA 300
    |||||
DB 241 QAERASSKSMITFDLKNKEVSVKRVTOPDKLQMGKLLPLHLTLPOALPOYAGSGNLTLA 300
QY 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKLMSTLKENKEAVSKKEKPVWV 360
    |||||
DB 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKLMSTLKENKEAVSKKEKPVWV 360
QY 361 LNPEAGMOCCLSDSGVLLSESNIKVLPWTSTPVHAPRALPAPTGSALDPQTASALP 420
    |||||
DB 361 LNPEAGMOCCLSDSGVLLSESNIKVLPWTSTPVHAPRALPAPTGSALDPQTASALP 420
QY 408 SALPDPQTASALPDPASALPAL 432
    |||||
DB 421 TALGLCLVSVFPPEPTVSWNSGAL 445

RESULT 5
AAP93008
ID AAP93008 standard; protein; 729 AA.
XX
AC AAP93008;
XX
DT 25-MAR-2003 (revised)
DT 02-NOV-1992 (first entry)
XX
DE Genetic construct which encodes CD4 linked to human IgG1 at the Hind3
    site upstream of the CH1 region (fusion protein CD4H-gamma-1).
XX
KW Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy;
    diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.
XX
OS Homo sapiens.
XX
PN EP325262-A.
XX
PD 26-JUL-1989.
XX
PF 20-JAN-1989; 89EP-00100913.
XX
PR 22-JAN-1988; 88US-00147351.
XX
PA (GENO ) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
DR WPI; 1989-214472/30.

```

DR N-PSDB; AAN90356.  
XX Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV  
PT infections or detecting HIV or SIV in sample.  
XX  
XX Example; Table 1, Page 12-23; 68pp; English.  
XX  
XX The fusion protein genes of the invention pref. comprise cDNA sequences  
CC which encode CD4 or a fragment which binds gp120 ligated to an expression  
CC plasmid which encodes an antibody in which the variable region of the  
CC gene has been deleted (see WO87-02671). The CD4 portion of the fusion  
CC protein may comprise the complete CD4 sequence, the 370 AA extracellular  
CC region and the membrane spanning domain, or the extracellular region. The  
CC Ig heavy chain is pref. from IgM, IgG1 or IgG3. The following are  
CC specifically claimed: fusion proteins CD4H-gamma-1, CD4Hmu, CD4E-  
CC gamma1, and CD4Hmu (No. 67608), pCD4F-gamma (No. 67609) and pCD4E-gamma-1  
CC (No. 67610). The plasmid containing (pCD4H-gamma-1) has been deposited in  
CC E. coli (MC1061/P3) at the ATCC under accession number 67611. (Updated on  
CC 25-MAR-2003 to correct PA field.)  
XX  
SQ Sequence 729 AA;  
  
Query Match 87.8%; Score 2042; DB 1; Length 729;  
Best Local Similarity 92.3%; Pred. No. 5, 6e-131;  
Matches 405; Conservative 3; Mismatches 9; Indels 22; Gaps 4;  
  
QY 1 MNRGVPFRHLVLTQALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNNOIK 60  
DB 1 MNRGVPFRHLVLTQALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNNOIK 60  
QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIKNLKIEDSDTYICEVEDQKEEYOL 120  
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIKNLKIEDSDTYICEVEDQKEEYOL 120  
QY 121 LVFGLTANSDFHLQGGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
DB 121 LVFGLTANSDFHLQGGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
QY 181 TWCTCTVLONOKKVEFKIDIVLAFQKASSIVYKKEGEVEFSPLAFTVEKLTSGSELMW 240  
DB 181 TWCTCTVLONOKKVEFKIDIVLAFQKASSIVYKKEGEVEFSPLAFTVEKLTSGSELMW 240  
QY 241 QAERASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOALPOYAGSGLTLA 300  
DB 241 QAERASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOALPOYAGSGLTLA 300  
QY 301 LEAKTGKLGHOEVNLYVMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWV 360  
DB 301 LEAKTGKLGHOEVNLYVMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWV 360  
QY 361 LNPEAGMOCCLSDSGQVLTLESNIKVLPTWSTPVH--PRA-----SALPAPPT-----G 407  
DB 361 LNPEAGMOCCLSDSGQVLTLESNIKVLPTWSTPVHADPEASTKGSPVFLAPSSKSTSGG 420  
QY 408 SA-----LPDPQTAS 417  
DB 421 TAAAGCLVVDYFPEPVTVS 439  
  
RESULT 6  
AAY51078  
ID AAY51078 standard; protein; 729 AA.  
XX  
XX AAY51078;  
XX  
XX 23-MAR-2000 (first entry)  
XX  
XX Human fusion protein CD4H-1.  
XX  
XX Fusion protein; human; CD4; IgG1; immunoglobulin; gp120;  
KM anti-human immunodeficiency virus; CD4H-1.  
OS Homo sapiens.

OS Synthetic.  
XX  
XX US6004781-A.  
XX  
XX 21-DEC-1999.  
XX  
XX 04-FEB-1994; 94US-00191708.  
XX  
XX 22-JAN-1988; 88US-00147351.  
XX 23-JAN-1989; 89US-00293596.  
XX 09-JUN-1992; 92US-00896781.  
XX 12-APR-1993; 93US-00057952.  
XX  
XX (GENO ) GEN HOSPITAL CORP.  
XX  
XX Seed B;  
XX  
XX WPI; 2000-085792/07.  
XX  
XX N-PSDB; AA244061.  
XX  
XX Fusion protein useful for the treatment of human immunodeficiency virus.  
XX  
XX Example 1; Col 15-30; 39pp; English.  
XX  
XX This invention describes a novel nucleic acid (I) encoding a fusion  
XX protein comprising a DNA sequence encoding amino acids 1-173 of CD4 (II)  
XX and a DNA sequence encoding a human immunoglobulin (Ig) heavy or light  
XX chain (III). The products of the invention have anti-human  
XX immunodeficiency virus (HIV) activity and are capable of binding to  
XX gp120. The fusion protein is useful for treating human immunodeficiency  
XX virus (HIV) or simian immunodeficiency virus (SIV). This sequence  
XX represents the fusion protein CD4H-1 which is constructed from CD4 linked  
XX to human IgG1 upstream of the CH1 region  
XX  
SQ Sequence 729 AA;  
  
Query Match 87.8%; Score 2042; DB 3; Length 729;  
Best Local Similarity 92.3%; Pred. No. 5, 6e-131;  
Matches 405; Conservative 3; Mismatches 9; Indels 22; Gaps 4;  
  
QY 1 MNRGVPFRHLVLTQALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNNOIK 60  
DB 1 MNRGVPFRHLVLTQALLPAATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNNOIK 60  
QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIKNLKIEDSDTYICEVEDQKEEYOL 120  
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIKNLKIEDSDTYICEVEDQKEEYOL 120  
QY 121 LVFGLTANSDFHLQGGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
DB 121 LVFGLTANSDFHLQGGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
QY 181 TWCTCTVLONOKKVEFKIDIVLAFQKASSIVYKKEGEVEFSPLAFTVEKLTSGSELMW 240  
DB 181 TWCTCTVLONOKKVEFKIDIVLAFQKASSIVYKKEGEVEFSPLAFTVEKLTSGSELMW 240  
QY 241 QAERASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOALPOYAGSGLTLA 300  
DB 241 QAERASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKPLHLTLPOALPOYAGSGLTLA 300  
QY 301 LEAKTGKLGHOEVNLYVMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWV 360  
DB 301 LEAKTGKLGHOEVNLYVMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVWV 360  
QY 361 LNPEAGMOCCLSDSGQVLTLESNIKVLPTWSTPVH--PRA-----SALPAPPT-----G 407  
DB 361 LNPEAGMOCCLSDSGQVLTLESNIKVLPTWSTPVHADPEASTKGSPVFLAPSSKSTSGG 420  
QY 408 SA-----LPDPQTAS 417  
DB 421 TAAAGCLVVDYFPEPVTVS 439

```

RESULT 7
ID AAY59168 standard; protein; 729 AA.
XX
AC AAY59168;
XX
DT 14-MAR-2000 (first entry)
XX
DE CD4-Ig fusion protein CD4Hgamma1.
XX
KM HIV, extracellular; CD4, gp120; immunoglobulin; Ig; fusion protein;
KM secreted protein; SIV infection; medicament.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN CA1340741-C.
XX
PD 14-SEP-1999.
XX
PF 20-JAN-1989; 89CA-00588749.
XX
PR 20-JAN-1989; 89CA-00588749.
XX
PA (GEMO ) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
PI WPI; 2000-063015/06.
DR N-PSDB; AA248201.
PT New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in
PT the treatment of HIV or simian immunodeficiency virus infections.
XX
PS Example 1; Page 25-36; 89pp; English.
XX
CC The invention provides a fusion gene encoding a fusion protein that
CC comprises an extracellular CD4 DNA sequence or its fragment which binds
CC to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA
CC sequence of an Ig heavy or light chain, where the DNA sequence encoding
CC the variable region has been replaced with the DNA sequence which encodes
CC extracellular CD4 or its gp120 binding fragment. The fusion protein is
CC capable of being secreted. The fusion proteins are useful for treating
CC HIV or SIV infections in animals, preferably humans. They are also useful
CC for producing medicaments which can be used for treating HIV or SIV
CC infections in humans. The present sequence represents the fusion protein
CC CD4Hgamma1 where the CD4 is linked to human IgG1 at the Hind3 site
CC upstream of the CHI region
CC
SQ Sequence 729 AA;
Query Match 87.8%; Score 2042; DB 3; Length 729;
Best Local Similarity 92.3%; Pred. No. 5,6e-131;
Matches 405; Conservative 3; Mismatches 9; Indels 22; Gaps 4;
QY 1 MNRGVFPRHLLVLVQLALLPATQGNKVVLGKGDVTELTCTASQKSIQPHMKNNSQIK 60
DB 1 MNRGVFPRHLLVLVQLALLPATQGNKVVLGKGDVTELTCTASQKSIQPHMKNNSQIK 60
QY 61 ILGNQGSFLTKGPSKLNDRADSRRLMDQGNFPLIIKKLTKEDSTTYICEVEDQKEEYQL 120
DB 61 ILGNQGSFLTKGPSKLNDRADSRRLMDQGNFPLIIKKLTKEDSTTYICEVEDQKEEYQL 120
QY 121 LVFGLTANSDPHLLQGGSLTTLTLESPGSSPSVOCRSRGRKNIOGKTLSTVSOLELDQSG 180
DB 121 LVFGLTANSDPHLLQGGSLTTLTLESPGSSPSVOCRSRGRKNIOGKTLSTVSOLELDQSG 180
QY 181 TWTCTVLONOKKVEFKIDIVLAFQKASSIYKKEGEVSEFPLAFTVEKLTGSGELMW 240
DB 181 TWTCTVLONOKKVEFKIDIVLAFQKASSIYKKEGEVSEFPLAFTVEKLTGSGELMW 240
QY 241 QAERASSSKSWITPPLKNKEVSVKRVTDPPKLGKSKLPLHLTLPQALPOYAGSGNLTLA 300
DB 241 QAERASSSKSWITPPLKNKEVSVKRVTDPPKLGKSKLPLHLTLPQALPOYAGSGNLTLA 300

```

```

DB 241 QAERASSSKSWITPDLKNKEVSVKRVTDPPKLGKSKLPLHLTLPQALPOYAGSGNLTLA 300
QY 301 LEAKTGKLEHENVLVVWRATQLOKNLTCEYMGPTSPKMLSLKLENBAKYSKREKPVWV 360
DB 301 LEAKTGKLEHENVLVVWRATQLOKNLTCEYMGPTSPKMLSLKLENBAKYSKREKPVWV 360
QY 361 LNPEAGMOCCLSDSGVLLSESNIKVLPTWSTPVH--PRA-----SALPAPPT-----G 407
DB 361 LNPEAGMOCCLSDSGVLLSESNIKVLPTWSTPVHADPEASTKGPVFPPLAPSSKSTSGG 420
QY 408 SA-----LPDPQTA 417
DB 421 TAAAGCLVKDYFPEPVTVS 439

RESULT 8
ID AAB19509 standard; protein; 416 AA.
XX
AC AAB19509;
XX
DT 09-JAN-2001 (first entry)
XX
DE CD4-IgM fusion protein CH4Mmu.
XX
KM CD4; IgM; human; CD4Mmu; fusion protein; immunoglobulin; HIV; SIV; gp120;
KM therapy; diagnosis.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH Protein 1..395
FT /note="CD4 extracellular region"
FT Protein 400..416
FT /note="IgM heavy chain partial sequence"
XX
PN US6117656-A.
XX
PD 12-SEP-2000.
XX
PF 07-JUN-1995; 95US-00479353.
XX
PR 22-JAN-1988; 88US-00147351.
PR 23-JAN-1989; 89US-00299596.
PR 09-JUN-1992; 92US-00896781.
PR 12-APR-1993; 93US-00057952.
PR 04-FEB-1994; 94US-00191708.
XX
PA (GEMO ) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
PI WPI; 2000-586558/55.
DR N-PSDB; AAA50662.
XX
CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or
SIV.
PS Example 1; Col 41-50; 39pp; English.
XX
CC The present sequence is that of fusion protein CD4Mmu comprising the
CC extracellular portion of CD4, which binds to HIV gp120, linked at its C-
CC terminus to the human IgM heavy chain. To obtain the fusion protein, DNA
CC encoding CD4 was linked to IgM DNA at the MspI site upstream of the CHI
CC region (see AAA50662). Fusion protein CD4Mmu and a nucleic acid encoding
CC it are claimed. Also claimed are a vector comprising the nucleic acid,
CC and a method of producing the fusion protein in secreted form using a
CC transformed host cell. The fusion protein may further comprise a
CC therapeutic agent, radiolabel or NMR imaging agent. The fusion protein
CC can be administered to an animal (including humans) for treatment of HIV
CC or SIV infection, and can also be used in assays for HIV or SIV, imaging
CC and tissue stains. IgM fusion proteins such as CD4Mmu provide complement-
CC mediated immunity.

```

XX Sequence 416 AA;  
SQ

Query Match 87.8%; Score 2040.5; DB 3; Length 416;  
Best Local Similarity 97.6%; Pred. No. 3, 6e-131;  
Matches 399; Conservative 0; Mismatches 3; Indels 7; Gaps 1;

QY 1 NMRGVPFRLHLVLTALPAPATQGNKVVLGKKGDTVELTCTASQKSIQFMKNSNOIK 60  
DB 1 NMRGVPFRLHLVLTALPAPATQGNKVVLGKKGDTVELTCTASQKSIQFMKNSNOIK 60  
QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIIKNLKIEDSDTYICEVEDQKEEYQL 120  
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIIKNLKIEDSDTYICEVEDQKEEYQL 120  
QY 121 LVFGILTANSDTHLQOGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
DB 121 LVFGILTANSDTHLQOGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
QY 181 TWTCTVLONOKKVEFKIDIVLAFOKASSIYKKGEQVEFSPLAFTVEKLTGSGELMW 240  
DB 181 TWTCTVLONOKKVEFKIDIVLAFOKASSIYKKGEQVEFSPLAFTVEKLTGSGELMW 240  
QY 241 QAERASSSKSWITFDLKNKEVSVKRVTPDKLQMGKULPLHLTLPOALPOYAGSGNLTIA 300  
DB 241 QAERASSSKSWITFDLKNKEVSVKRVTPDKLQMGKULPLHLTLPOALPOYAGSGNLTIA 300  
QY 301 LEAKTGKLIHOEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKENKAQVSKREKPVWV 360  
DB 301 LEAKTGKLIHOEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKENKAQVSKREKPVWV 360  
QY 361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTPVHPRASALPAPPTGSA 409  
DB 361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTPVH-----ADPGSA 402

RESULT 9  
AAR27277

ID AAR27277 standard; protein; 462 AA.

XX AAR27277;  
XX 25-MAR-2003 (revised)  
DT 28-JUL-1995 (first entry)  
XX  
XX CD4:eta peptide chimeric protein.  
XX  
XX Fusion protein; CD4; extracellular domain; zeta; eta; gamma;  
XX membrane spanning domain; intracellular domain; type 1;  
XX integral membrane homodimer; TCR; T cell antigen receptor;  
XX extracellular domain; mouse; human; receptor; chimera;  
XX HPB-ALL tumour cell line; natural killer cell.  
XX  
XX Homo sapiens.

XX Key Location/Qualifiers  
FH 1..399  
FT /note="CD4 extracellular domain"  
FT Protein 400..462  
FT /note="Zeta membrane spanning and intracellular domain"  
XX  
XX MO9215322-AL.

XX PD 17-SEP-1992.  
XX  
XX 06-MAR-1992; 92WO-US001785.  
XX  
XX 07-MAR-1991; 91US-00665961.  
XX  
XX (GEHO ) GEN HOSPITAL CORP.  
XX  
XX Seed B, Romeo C, Kolanus W;  
XX

DR WPI; 1992-331474/40.  
DR N-PSDB; AAQ28705.

PT Therapeutic cells expressing chimeric receptors - directing cellular  
PT response to an infective agent; useful in treating HIV-1, AIDS  
PT Pneumocystis carinii infections etc.

PS Example 2; Page 73-74; 11app; English.

XX This sequence represents a fusion protein between the CD4 extracellular  
XX domain and the eta protein membrane spanning domain and intracellular  
XX domain. Eta is an isoform of zeta (see also AAR27276) which is a 32 kd  
XX type I integral membrane homodimer, which arises by alternate mRNA  
XX splicing. It is present in reduced amounts in cells expressing the T cell  
XX antigen receptor. Zeta-eta heterodimers are thought to mediate the  
XX formation of inositol phosphates, as well as the receptor initiated cell  
XX death called apoptosis. In the production of the CD4 receptor chimera,  
XX the eta CDNA was isolated from the HPB-ALL tumour cell line and from  
XX human natural killer cells. The eta CDNA was joined to the extracellular  
XX domain of an engineered form of CD4 possessing a BamHI site just upstream  
XX of the membrane spanning domain, by a BamHI site naturally present a few  
XX residues upstream of the membrane spanning domain. (Updated on 25-MAR-  
XX 2003 to correct PN field.)

SQ Sequence 462 AA;

Query Match 87.7%; Score 2040; DB 2; Length 462;  
Best Local Similarity 91.4%; Pred. No. 4, 4e-131;  
Matches 405; Conservative 1; Mismatches 9; Indels 28; Gaps 2;

QY 1 NMRGVPFRLHLVLTALPAPATQGNKVVLGKKGDTVELTCTASQKSIQFMKNSNOIK 60  
DB 1 NMRGVPFRLHLVLTALPAPATQGNKVVLGKKGDTVELTCTASQKSIQFMKNSNOIK 60  
QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIIKNLKIEDSDTYICEVEDQKEEYQL 120  
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIIKNLKIEDSDTYICEVEDQKEEYQL 120  
QY 121 LVFGILTANSDTHLQOGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
DB 121 LVFGILTANSDTHLQOGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
QY 181 TWTCTVLONOKKVEFKIDIVLAFOKASSIYKKGEQVEFSPLAFTVEKLTGSGELMW 240  
DB 181 TWTCTVLONOKKVEFKIDIVLAFOKASSIYKKGEQVEFSPLAFTVEKLTGSGELMW 240  
QY 241 QAERASSSKSWITFDLKNKEVSVKRVTPDKLQMGKULPLHLTLPOALPOYAGSGNLTIA 300  
DB 241 QAERASSSKSWITFDLKNKEVSVKRVTPDKLQMGKULPLHLTLPOALPOYAGSGNLTIA 300  
QY 301 LEAKTGKLIHOEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKENKAQVSKREKPVWV 360  
DB 301 LEAKTGKLIHOEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKENKAQVSKREKPVWV 360  
QY 361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTPVHPRASALPAPPTGSA 420  
DB 361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTPVH-----ADPQ----- 399  
QY 421 DPPASALPALAVISFLGLGL 443  
DB 400 -----LCYILDALFLYGIIVL 415

RESULT 10  
AAR78677

ID AAR78677 standard; protein; 462 AA.

XX AAR78677;  
XX  
XX 16-APR-1996 (first entry)  
XX  
XX T-cell receptor gamma.  
XX  
XX





```

Db      181 TWTCVTLQNKQKVEFKIDIVLAFQKASSIYVKKGEQVEFSFPPLAFTVEKLTGSGELMW 240
Qy      241 QAERASSSSKSWITFDLKNKEVSVKRYTODPKLQMGKULPLHLTLPOALPOVAGSNLTLA 300
Db      241 QAERASSSSKSWITFDLKNKEVSVKRYTODPKLQMGKULPLHLTLPOALPOVAGSNLTLA 300
Qy      301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKYSKREKPVWV 360
Db      301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKYSKREKPVWV 360
Qy      361 LNPEAGMWQCLLSDSGQVLLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
Db      361 LNPEAGMWQCLLSDSGQVLLLESNIKVLPTWSTPVH-----ADPQ----- 399
Qy      421 DPPAASALPALAVISFLIGLGL 443
Db      400 -----LCYILDALFLFYGLV 415

RESULT 12
AAW02214 ID AAW02214 standard; protein; 462 AA.
AAW02214;
AC AAW02214;
XX
XX
XX 11-NOV-1996 (first entry)
DT
XX
XX CD4:Fc receptor gamma chain chimaeric receptor.
DE
XX
XX Chimaeric receptor; cellular immunity; adoptive immunotherapy; CD4;
KM human immunodeficiency virus type 1; HIV-1; AIDS; therapy;
KW Fc receptor gamma chain; cytotoxic T lymphocyte; CTL.
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
FH 1. 393
FT /label= "Extracellular domain"
FT /note= "CD4 extracellular domain"
FT
FT Region 394. 397
FT /label= linker
FT /note= "encoding DNA contains a BamHI site used for
FT fusion construction"
FT 398. 462
FT /note= "region of fusion derived from gamma chain,
FT preferred signal-transducing portions for constructs of
FT the invention are amino acids 421-462 and 402-419"
FT 400. 462
FT /label= Transmembrane+intracellular_domains
FT
FT Domain
XX
XX W09625953-A1.
XX
XX 29-AUG-1996.
XX
XX 25-JAN-1996; 96WO-US001056.
XX
XX 24-FEB-1995; 95US-00394176.
XX
XX (GEHO ) GEN HOSPITAL CORP.
XX
XX Seed B, Romeo C, Kolanus W;
XX
XX WPI. 1996-402134/40.
XX
XX N-PSDB; AAT36759.
XX
XX
XX Direction of cellular immune response using therapeutic cell expressing 2
XX chimaeric receptors - comprising region binding to target cell and region
XX that signals target cell destruction, or CD28 region, partic. for
XX eliminating HIV-infected cells.
XX
XX Claim 7, Page 76; 120P; English.
XX
XX A chimaeric receptor (AAW0214) comprises the extracellular domain of an

```

```

CC engineered form of the CD4 cellular receptor for HIV and the
CC transmembrane and intracellular regions, including the cytoytic signal-
CC transducing portion, of the human Fc receptor gamma chain; the region of
CC the fusion is shown in AAW02223. It can be obtd. by inserting a gene
CC fusion (AAT36759) into a vaccinia virus vector and expression in host
CC cells. Chimaeric receptors comprising CD4 fused to Fc receptor gamma or T
CC -cell receptor zeta (see also AAW02213) or eta (AAW02215) chains are
CC capable of directing cytotoxic T lymphocytes to specifically recognise
CC and kill cells expressing HIV gp120, thus providing a therapy for AIDS
XX
XX Sequence 462 AA;
SQ
Query Match 87.7%; Score 2040; DB 2; Length 462;
Best Local Similarity 91.4%; Pred. No. 4,4e-131;
Matches 405; Conservative 1; Mismatches 9; Indels 28; Gaps 2;
Qy 1 MNRGVFPHLLLVLTQALLPAAQGNKRVLLGKGDVLELTTSQKKSIOFHMKNNOIK 60
Db 1 MNRGVFPHLLLVLTQALLPAAQGNKRVLLGKGDVLELTTSQKKSIOFHMKNNOIK 60
Qy 61 ILNGGSFLTKGSPKLNDRADSRSLMDQGNPPLIKLKTIEDSTYICEVEDQKEVQL 120
Db 61 ILNGGSFLTKGSPKLNDRADSRSLMDQGNPPLIKLKTIEDSTYICEVEDQKEVQL 120
Qy 121 LVFGLTANSDPHLLQGGSLTLTLSPGSSPVOCSPRGKNIQGGKTLVSQLELDQSG 180
Db 121 LVFGLTANSDPHLLQGGSLTLTLSPGSSPVOCSPRGKNIQGGKTLVSQLELDQSG 180
Qy 181 TWTCVTLQNKQKVEFKIDIVLAFQKASSIYVKKGEQVEFSFPPLAFTVEKLTGSGELMW 240
Db 181 TWTCVTLQNKQKVEFKIDIVLAFQKASSIYVKKGEQVEFSFPPLAFTVEKLTGSGELMW 240
Qy 241 QAERASSSSKSWITFDLKNKEVSVKRYTODPKLQMGKULPLHLTLPOALPOVAGSNLTLA 300
Db 241 QAERASSSSKSWITFDLKNKEVSVKRYTODPKLQMGKULPLHLTLPOALPOVAGSNLTLA 300
Qy 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKYSKREKPVWV 360
Db 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKYSKREKPVWV 360
Qy 361 LNPEAGMWQCLLSDSGQVLLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
Db 361 LNPEAGMWQCLLSDSGQVLLLESNIKVLPTWSTPVH-----ADPQ----- 399
Qy 421 DPPAASALPALAVISFLIGLGL 443
Db 400 -----LCYILDALFLFYGLV 415

RESULT 13
AAW83142 ID AAW83142 standard; protein; 462 AA.
AAW83142;
AC AAW83142;
XX
XX 03-FEB-1999 (first entry)
DT
XX
XX Chimeric receptor containing mouse gamma polypeptide.
DE
XX
XX Human; zeta; eta; gamma; membrane-bound chimeric receptor; infection;
XX tumour; cancer cell; autoimmune-generated cell; T cell receptor; CD3;
XX CD4; B cell receptor; Fc receptor; pathogen; bacterial; fungal;
XX protozoan; viral.
XX
XX OS Synthetic.
XX
XX Mus sp.
XX
XX US5643728-A.
XX
XX 01-DEC-1998.
XX
XX 05-APR-1995; 95US-00417495.
XX

```

PR 07-MAR-1991; 91US-00665961.  
 PR 06-MAR-1992; 92US-00847566.  
 PR 28-FEB-1994; 94US-00203866.  
 XX  
 PA (GENO ) GEN HOSPITAL CORP.  
 XX  
 PI Romeo C, Kolanus W, Seed B;  
 XX  
 DR WPI; 1999-044582/04.  
 DR N-PSDB; AAV70158.  
 XX  
 PT Membrane-bound chimeric receptors - comprising extracellular portion  
 PT which recognises and binds a target cell and an intracellular portion of  
 PT e.g. a T-cell receptor.  
 XX  
 PS Example 2; Col 43-46; 57bp; English.  
 XX  
 CC The present invention describes DNA encoding a membrane-bound chimeric  
 CC receptor comprising: (a) an extracellular portion that specifically  
 CC recognises and binds a target cell or a target infective agent; and (b)  
 CC an intracellular portion of a T-cell receptor CD3, zeta or eta  
 CC polypeptide, a B-cell receptor polypeptide or an Fc receptor polypeptide.  
 CC The present sequence represents a chimeric receptor containing the mouse  
 CC gamma polypeptide. Cells expressing chimeric receptors of the present  
 CC invention can be administered to mammals in order to destroy pathogens  
 CC (e.g. bacteria, fungi, protozoa or viruses, especially HIV), cancer cells  
 CC or autoimmune-generated cells  
 XX  
 SQ Sequence 462. AA;

Query Match 87.7%; Score 2040; DB 2; Length 462;  
 Best Local Similarity 91.4%; Pred. No. 4.4e-131;  
 Matches 405; Conservative 1; Mismatches 9; Indels 28; Gaps 2;

QY 1 MNRGVPFRHLLVLTQALLPAAATQGNKVYLGKGGDTVELTCTASQKSIQFHMKNNSQIK 60  
 DB 1 MNRGVPFRHLLVLTQALLPAAATQGNKVYLGKGGDTVELTCTASQKSIQFHMKNNSQIK 60  
 QY 61 ILNGGSEFLTKGSPKLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVOL 120  
 DB 61 ILNGGSEFLTKGSPKLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVOL 120  
 QY 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDQSG 180  
 QY 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDQSG 180  
 QY 181 TWTCTVLONQKKEFKIDIVVLAFOKASSIVYKKEGEQVEFSFPLATVEKLTGSGELMW 240  
 DB 181 TWTCTVLONQKKEFKIDIVVLAFOKASSIVYKKEGEQVEFSFPLATVEKLTGSGELMW 240  
 QY 241 QAEKASSKSWITTFPLKKNKEVSVKAVTODPKLQMGKPLHLTLTPQALPOYAGSGLTLA 300  
 DB 241 QAEKASSKSWITTFPLKKNKEVSVKAVTODPKLQMGKPLHLTLTPQALPOYAGSGLTLA 300  
 QY 301 LEAKTGKLGHEVNLVVMRAATQLOKNLTCCEVWGPTSPKMLSLIKENKEAKVSKREKPVW 360  
 DB 301 LEAKTGKLGHEVNLVVMRAATQLOKNLTCCEVWGPTSPKMLSLIKENKEAKVSKREKPVW 360  
 QY 361 LNPEKGMQOCLLSDSGOVLLEBNINIVLPTWSTPVHPRAALPAAPTGSALDPPQASALP 420  
 DB 361 LNPEKGMQOCLLSDSGOVLLEBNINIVLPTWSTPVH-----ADPQ----- 399  
 QY 421 DPPASALPAALAVISFLGLGL 443  
 DB 400 -----LCYILDALFLXGYVL 415

RESULT 14  
 ID AAR27276 standard; proteoin; 575 AA.  
 AC AAR27276;  
 XX

DT 25-MAR-2003 (revised)  
 DT 28-JUL-1995 (first entry)  
 XX  
 DE CD4:zeta peptide chimeric protein.  
 XX  
 KW Fusion protein; CD4; extracellular domain; zeta; eta; gamma;  
 KW membrane spanning domain; intracellular domain; type I;  
 KW integral membrane homodimer; TCR; T cell antigen receptor;  
 KW extracellular domain; mouse; human; receptor; chimera;  
 KW HbB-ALL tumour cell line; natural killer cell.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key  
 FT Protein 1..399  
 FT /note= "CD4 extracellular domain"  
 FT Protein 400..575  
 FT /note= "Zeta membrane spanning and intracellular domain"  
 XX  
 PN MO9215322-A1.  
 XX  
 PD 17-SEP-1992.  
 XX  
 XX 06-MAR-1992; 92WO-US001785.  
 PF  
 XX 07-MAR-1991; 91US-00665961.  
 PR  
 XX (GENO ) GEN HOSPITAL CORP.  
 PA  
 PI Seed B, Romeo C, Kolanus W;  
 XX  
 XX WPI; 1992-331474/40.  
 DR N-PSDB; AAQ28704.  
 DR  
 XX  
 PT Therapeutic cells expressing chimeric receptors - directing cellular  
 PT response to an infective agent, useful in treating HIV-1, AIDS  
 PT Pneumocystis carinii infections etc.  
 XX  
 PS Example 2; Page 72-73; 114bp; English.  
 XX  
 CC This sequence represents a fusion protein between the CD4 extracellular  
 CC domain and the zeta protein membrane spanning domain and intracellular  
 CC domain. Zeta is a 32 kD type I integral membrane homodimer which has a 9  
 CC residue extracellular domain and a 112/113 residue intracellular domain  
 CC for mouse and human protein respectively. In the production of the CD4  
 CC receptor chimera, the zeta cDNA was isolated from the HbB-ALL tumour cell  
 CC line and from human natural killer cells. The zeta cDNA was joined to the  
 CC extracellular domain of an engineered form of CD4 possessing a BamHI site  
 CC just upstream of the membrane spanning domain, by a BamHI site naturally  
 CC present a few residues upstream of the membrane spanning domain. (updated  
 CC on 25-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 575 AA;

Query Match 87.7%; Score 2038.5; DB 2; Length 575;  
 Best Local Similarity 92.1%; Pred. No. 7.3e-131;  
 Matches 406; Conservative 4; Mismatches 8; Indels 23; Gaps 3;  
 QY 1 MNRGVPFRHLLVLTQALLPAAATQGNKVYLGKGGDTVELTCTASQKSIQFHMKNNSQIK 60  
 DB 1 MNRGVPFRHLLVLTQALLPAAATQGNKVYLGKGGDTVELTCTASQKSIQFHMKNNSQIK 60  
 QY 61 ILNGGSEFLTKGSPKLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVOL 120  
 DB 61 ILNGGSEFLTKGSPKLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVOL 120  
 QY 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDQSG 180  
 QY 181 TWTCTVLONQKKEFKIDIVVLAFOKASSIVYKKEGEQVEFSFPLATVEKLTGSGELMW 240  
 DB 181 TWTCTVLONQKKEFKIDIVVLAFOKASSIVYKKEGEQVEFSFPLATVEKLTGSGELMW 240

```

QY 241 QAEKSSSSKSWITTPDLKKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGLTLA 300
DB 241 QAEKSSSSKSWITTPDLKKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGLTLA 300
QY 301 LEAKTGKHLQEVNLYVMRATQLOKNTLCEWGPSTSPKMLSLKLENKEAKVSKREKPVWV 360
DB 301 LEAKTGKHLQEVNLYVMRATQLOKNTLCEWGPSTSPKMLSLKLENKEAKVSKREKPVWV 360
QY 361 LNPEAGMOCCLSDSGQVLLSNIKVLPTWSTPVH--PR-----ASAL 401
DB 361 LNPEAGMOCCLSDSGQVLLSNIKVLPTWSTPVHADPKLCYLLDGLIFIVGIITLALYL 420
QY 402 PAPPTGSALPDPQTASALPDP 422
DB 421 RAKFSRSA-----ETANLQDP 437

RESULT 15
AAR78676
ID AAR78676 standard; protein; 575 AA.
AC AAR78676;
XX
XX 16-APR-1996 (first entry)
DT
XX
XX T-cell receptor zeta.
DE
XX
XX Chimeric receptor; CD4; T-cell receptor zeta; HIV; cytolysis;
KW human immunodeficiency virus; adoptive immunotherapy.
XX
XX Homo sapiens.
OS
XX
XX MO9521528-A1.
PN
XX
XX 17-AUG-1995.
PD
XX
XX 12-JAN-1995; 95WO-US000454.
PF
XX
XX 14-FEB-1994; 94US-00195395.
PR 02-AUG-1994; 94US-00284391.
XX
XX (GEMO ) GEN HOSPITAL CORP.
PA
XX
XX Seed B, Banapour B, Romeo C, Kolanus W;
PI
XX
XX WPI; 1995-292893/38.
DR N-PSDB; AAQ96122.
XX
XX Target cytolysis of HIV-infected cells - by chimeric CD4 receptor-bearing
PT cells.
XX
XX Example 2; Page 76-77; 118pp; English.
PS
XX
XX Fusion proteins comprising the extracellular domain of CD4 fused to T-
CC cell receptor zeta, gamma or eta (AAR78676-78, respectively) were
CC expressed in Cyt1 using a vaccine virus vector. These CD4:zeta, CD4:gamma
CC and CD4:eta chimeric receptors mediated cytolysis of targets expressing
CC HIV gp120/41
CC
XX Sequence 575 AA;
SQ

Query Match 87.7%; Score 2038.5; DB 2; Length 575;
Best Local Similarity 92.1%; Pred. No. 7.3e-131;
Matches 406; Conservative 4; Mismatches 8; Indels 23; Gaps 3;

```

```

QY 121 LVFGLTANSDPHLLOQGSITLTLSPSPSSPVOCRSRGRKNIOGKTLVSQLELQDSG 180
DB 121 LVFGLTANSDPHLLOQGSITLTLSPSPSSPVOCRSRGRKNIOGKTLVSQLELQDSG 180
QY 181 TWCTCTVLQNOKKVEKIDIVVLAFOKASSIYKKEGQVERSFPLAFTVEKLTSGGELMW 240
DB 181 TWCTCTVLQNOKKVEKIDIVVLAFOKASSIYKKEGQVERSFPLAFTVEKLTSGGELMW 240
QY 241 QAEKSSSSKSWITTPDLKKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGLTLA 300
DB 241 QAEKSSSSKSWITTPDLKKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGLTLA 300
QY 301 LEAKTGKHLQEVNLYVMRATQLOKNTLCEWGPSTSPKMLSLKLENKEAKVSKREKPVWV 360
DB 301 LEAKTGKHLQEVNLYVMRATQLOKNTLCEWGPSTSPKMLSLKLENKEAKVSKREKPVWV 360
QY 361 LNPEAGMOCCLSDSGQVLLSNIKVLPTWSTPVH--PR-----ASAL 401
DB 361 LNPEAGMOCCLSDSGQVLLSNIKVLPTWSTPVHADPKLCYLLDGLIFIVGIITLALYL 420
QY 402 PAPPTGSALPDPQTASALPDP 422
DB 421 RAKFSRSA-----ETANLQDP 437

RESULT 16
AAR89456
ID AAR89456 standard; protein; 575 AA.
AC AAR89456;
XX
XX 26-SEP-1996 (first entry)
DT
XX
XX CD4:zeta fusion protein.
DE
XX
XX CD7; transmembrane domain; chimeric receptor; CD5; CD34; CH2; CH3; IgG1;
KW human; CD4; HIV; proteinaceous alpha-helix; T cell; B cell; neutrophil;
KW dendritic cell; therapy; mammal; infection.
XX
XX Synthetic.
OS
XX
XX MO9603883-A1.
PN
XX
XX 15-FEB-1996.
PD
XX
XX 26-JUL-1995; 95WO-US009468.
PF
XX
XX 02-AUG-1994; 94US-00284391.
PR 24-FEB-1995; 95US-00394388.
XX
XX (GEMO ) GEN HOSPITAL CORP.
PA
XX
XX Seed B, Banapour B, Romeo C, Kolanus W;
PI
XX
XX WPI; 1996-129034/13.
DR N-PSDB; AAT10801.
XX
XX Membrane-bound chimeric receptor comprising extracellular portion
PT including CD4 fragment - cells expressing receptor can be used for
PT treatment of HIV infection.
XX
XX Example 2; Page 77-78; 134pp; English.
PS
XX
XX AAT10801-T10803 represent membrane bound proteinaceous chimeric receptors
CC of the invention. This sequence represents the CD4:zeta chimera. The
CC transmembrane region of the chimeric receptor acts to separate the
CC intracellular and extracellular domains of the chimera, and contains a
CC portion of the CD7 (see AAR89440), CD5 or CD34 transmembrane domains.
CC Alternatively, the extracellular portion of the receptor can be separated
CC from the intracellular domain by the hinge, CH2 and CH3 domains of human
CC IgG1 (see AAR89441). The extracellular portion of the chimeric receptor
CC contains a fragment of CD4 (amino acids 1-194 or 1-200 of the CD4

```

CC sequence, see AAR89450 and AAR89451) which specifically recognises and  
 CC binds HIV-infected cells, but does not mediate HIV infection. The  
 CC extracellular domain of the receptor is separated from the cell membrane  
 CC by 48 or 72 angstroms, or by one or more proteinaceous alpha-helices. The  
 CC cells expressing the receptor are preferably T cells, B cells,  
 CC neutrophils, or dendritic cells. The therapeutic cells expressing the  
 CC chimeric receptor are administered to a mammal to treat HIV infection  
 XX  
 SQ Sequence 575 AA;

Query Match 87.7%; Score 2038.5; DB 2; Length 575;  
 Best Local Similarity 92.1%; Pred. No. 7.3e-131;  
 Matches 406; Conservative 4; Mismatches 8; Indels 23; Gaps 3;

QY 1 MNRGVPFRHLVLTQALPLPATQGNKVLGKGGDTVELCTASQKSIQFHMKNNOIK 60  
 Db 1 MNRGVPFRHLVLTQALPLPATQGNKVLGKGGDTVELCTASQKSIQFHMKNNOIK 60  
 QY 61 ILNGGSFLTKGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQL 120  
 Db 61 ILNGGSFLTKGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQL 120  
 QY 121 LVFGILTANSDTHLQGGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 Db 121 LVFGILTANSDTHLQGGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 QY 181 TWCTVLOKQKVEFKIDIVLAFQKASSIYKKEGEQVFSFPLATFVEKLTGSGELMW 240  
 Db 181 TWCTVLOKQKVEFKIDIVLAFQKASSIYKKEGEQVFSFPLATFVEKLTGSGELMW 240  
 QY 241 QABRASSSKSWITFDLNKKEVSVKRVTDQPKLQMGKPLHLTLPLQALPOYAGSGNLTLA 300  
 Db 241 QABRASSSKSWITFDLNKKEVSVKRVTDQPKLQMGKPLHLTLPLQALPOYAGSGNLTLA 300  
 QY 301 LEAKTGKLEHGVNLVVMRAATQLOKRLTCEWGPSPKMLSLKLENKAKVSKKEKPVW 360  
 Db 301 LEAKTGKLEHGVNLVVMRAATQLOKRLTCEWGPSPKMLSLKLENKAKVSKKEKPVW 360  
 QY 361 LNPEAGMOCCLDSGQVLESINIVLPTWSTPVH--PR-----ASAL 401  
 Db 361 LNPEAGMOCCLDSGQVLESINIVLPTWSTPVHADPKLCYLLDGLIFIVGITLALYL 420  
 QY 402 PAPPTGSALPDPQTASALPDP 422  
 Db 421 RAKFSRSA----ETRAVLQDP 437

RESULT 17

AAW02213 standard; protein; 575 AA.

AAW02213;

11-NOV-1996 (first entry)

CD4: T-cell receptor zeta chain chimaeric receptor.

XX Chimaeric receptor: cellular immunity; adoptive immunotherapy; CD4;  
 KW human immunodeficiency virus type 1; HIV-1; AIDS; therapy;  
 KW T-cell receptor zeta chain; cytotoxic T lymphocyte; CTL.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Domain 1..393

FT /label= Extracellular\_domain

FT /note= "CD4 extracellular domain"

FT Region 394..396

FT /label= Linker

FT /note= "encoding DNA contains a BamHI site used for fusion construction"

FT Region 397..575

FT /note= "region of fusion derived from zeta chain,"

FT preferred signal-transducing portions for constructs of  
 FT the invention are amino acids 421-575, 423-255, 438-455,  
 FT 461-494, 494-528, 400-420 and 421-462"  
 FT 400..437  
 FT /label= Transmembrane domain  
 FT /note= "zeta chain transmembrane domain"  
 FT 438..575  
 FT /label= Intracellular domain  
 FT /note= "zeta chain intracellular domain"

XX W09625953-A1.

XX 29-AUG-1996.

XX 25-JAN-1996; 96WO-US001056.

XX 24-FEB-1995; 95US-00394176.

XX (GEHO ) GEN HOSPITAL CORP.

XX Seed B, Romeo C, Kolanus W;

XX WPI; 1996-402134/40.

XX N-PSDB; AAT36758.

PT Direction of cellular immune response using therapeutic cell expressing 2  
 PT chimaeric receptors - comprising region binding to target cell and region  
 PT that signals target cell destruction, or CD28 region, partic. for  
 PT eliminating HIV-infected cells.

XX Claim 7; Page 74-75; 120pp; English.

XX A chimaeric receptor (AAW02213) comprises the extracellular domain of an  
 CC engineered form of the CD4 cellular receptor for HIV and the  
 CC transmembrane and intracellular regions, including the cytosolic signal-  
 CC transducing portion, of the human T-cell receptor zeta chain; the region  
 CC of the fusion is shown in AAW02213. It can be obtd. by inserting a gene  
 CC fusion (AAT36758) into a vaccinia virus vector and expression in host  
 CC cells. Chimaeric receptors comprising CD4 fused to zeta, eta (see also  
 CC AAW02215) or Fc receptor gamma (see also AAW02214) chains are capable of  
 CC directing cytotoxic T lymphocytes to specifically recognise and kill  
 CC cells expressing HIV gp120, thus providing a therapy for AIDS

XX Sequence 575 AA;

Query Match 87.7%; Score 2038.5; DB 2; Length 575;  
 Best Local Similarity 92.1%; Pred. No. 7.3e-131;  
 Matches 406; Conservative 4; Mismatches 8; Indels 23; Gaps 3;

QY 1 MNRGVPFRHLVLTQALPLPATQGNKVLGKGGDTVELCTASQKSIQFHMKNNOIK 60

Db 1 MNRGVPFRHLVLTQALPLPATQGNKVLGKGGDTVELCTASQKSIQFHMKNNOIK 60

QY 61 ILNGGSFLTKGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQL 120

Db 61 ILNGGSFLTKGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQL 120

QY 121 LVFGILTANSDTHLQGGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180

Db 121 LVFGILTANSDTHLQGGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180

QY 181 TWCTVLOKQKVEFKIDIVLAFQKASSIYKKEGEQVFSFPLATFVEKLTGSGELMW 240

Db 181 TWCTVLOKQKVEFKIDIVLAFQKASSIYKKEGEQVFSFPLATFVEKLTGSGELMW 240

QY 241 QABRASSSKSWITFDLNKKEVSVKRVTDQPKLQMGKPLHLTLPLQALPOYAGSGNLTLA 300

Db 241 QABRASSSKSWITFDLNKKEVSVKRVTDQPKLQMGKPLHLTLPLQALPOYAGSGNLTLA 300

QY 301 LEAKTGKLEHGVNLVVMRAATQLOKRLTCEWGPSPKMLSLKLENKAKVSKKEKPVW 360

Db 301 LEAKTGKLEHGVNLVVMRAATQLOKRLTCEWGPSPKMLSLKLENKAKVSKKEKPVW 360

Qy	361	LNPEAGMMQCLLSDSGQVLLSESNIKYLPWSPVPH--PR-----ASL	401
Db	361	LNPEAGMMQCLLSDSGQVLLSESNIKYLPWSPVPHADPFLCYLLDILFRTGYITATLYL	420
Qy	402	PAPPTGSALPDQPTASALPDP	422
Db	421	RAKFSRSA---ETANLQDP	437
RESULT 18			
ID	AAW83140	standard; protein; 575 AA.	
XX	AAW83140;		
XX	03-FEB-1999	(first entry)	
XX	03-FEB-1999	(first entry)	
XX	Chimeric receptor containing human zeta polypeptide.		
XX	Human; zeta; eta; gamma; membrane-bound chimeric receptor; infection;		
XX	tumour; cancer cell; autoimmune-generated cell; T cell receptor; CD3;		
XX	CD4; B cell receptor; Fc receptor; pathogen; bacterial; fungal;		
XX	protozoan; viral.		
XX	Synthetic.		
XX	Homo sapiens.		
XX	US5843728-A.		
XX	01-DEC-1998.		
XX	05-APR-1995;	95US-00417495.	
XX	07-MAR-1991;	91US-00665961.	
XX	06-MAR-1993;	92US-00847566.	
XX	28-FEB-1994;	94US-00203866.	
XX	(GENO ) GEN HOSPITAL CORP.		
XX	Romeo C, Kolanus W, Seed B;		
XX	WPI; 1999-044582/04.		
XX	N-PSDB; AAV70156.		
XX	Membrane-bound chimeric receptors - comprising extracellular portion of		
XX	which recognises and binds a target cell and an intracellular portion of		
XX	e.g. a T-cell receptor.		
XX	Example 2; Col 39-42; 57pp; English.		
XX	The present invention describes DNA encoding a membrane-bound chimeric		
XX	receptor comprising: (a) an extracellular portion that specifically		
XX	recognises and binds a target cell or a target infective agent; and (b)		
XX	an intracellular portion of a T-cell receptor CD3, zeta or eta		
XX	polypeptide, a B-cell receptor polypeptide or an Fc receptor polypeptide.		
XX	The present sequence represents a chimeric receptor containing the human		
XX	zeta polypeptide. Cells expressing chimeric receptors of the present		
XX	invention can be administered to mammals in order to destroy pathogens		
XX	(e.g. bacteria, fungi, protozoa or viruses, especially HIV), cancer cells		
XX	or autoimmune-generated cells		
XX	Sequence 575 AA;		
Qy	Query Match	87.7%; Score 2038.5; DB 2; Length 575;	
Db	Best Local Similarity	92.1%; Pred. No. 7.3e-131;	
Qy	Matches	406; Conservative 4; Mismatches 8; Indels 23; Gaps 3	
Db	1	MNRGVPFRLLLVLTQALPPATQGNKVLGGKGDVETCTPASQKKSIOFHKNSNQK	60
Db	1	MNRGVPFRLLLVLTQALPPATQGNKVLGGKGDVETCTPASQKKSIOFHKNSNQK	60
Qy	1	ILVNGQSFLTKGPSKLNDRADRSRLMDQGNPFLIKNLKIEDSDTYICEVEDQKEVOL	120

Db	61	ILGNQSSFLTKGSSKLNDAADSRSLMDQGNFPLIIINKLKIEDSDTYICEVEDQKEVOL	120		
Qy	121	LVFGLTANSTPHLLQGSQSLTLTLSPSPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180		
Db	121	LVFGLTANSTPHLLQGSQSLTLTLSPSPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSG	180		
Qy	181	TWTCYVLQNKQKVEFKIDIVLAFQKASSIYKKKEGEVESPFLAFTVEKLTSSGELMW	240		
Db	181	TWTCYVLQNKQKVEFKIDIVLAFQKASSIYKKKEGEVESPFLAFTVEKLTSSGELMW	240		
Qy	241	QAEBASSSSKWTIFEDLKNKEVSKVVTOPPKLQMGKTLPLHLTLPOLAPQYAGSGNLTLA	300		
Db	241	QAEBASSSSKWTIFEDLKNKEVSKVVTOPPKLQMGKTLPLHLTLPOLAPQYAGSGNLTLA	300		
Qy	301	LEAKTGKLNHEVNLVVMRATOLQKNLTCEVWGPTSPKLMLSLKLENKAKVSKREKPVWV	360		
Db	301	LEAKTGKLNHEVNLVVMRATOLQKNLTCEVWGPTSPKLMLSLKLENKAKVSKREKPVWV	360		
Qy	361	LNPEAGMGMQCLSDSGQVLLSESNIKYLPWSTPVH--PR-----ASAL 401	420		
Qy	402	PAPPTGALPDPQTASALPDP 422			
Db	421	RAKFSRSA---ETAANLQDP 437			
RESULT 19					
ID	AAp81990	AAp81990 standard; protein; 458 AA.			
XX	AAp81990;				
AC	25-MAR-2003 (revised)				
XX	DT 15-OCT-1990 (first entry)				
DE	Clone pT4B encoded HIV T4 glycoprotein.				
XX					
KW	Human Immunodeficiency Virus T4 envelope glycoprotein; AIDS; ss.				
XX					
OS	Synthetic.				
XX					
XX	Key	Location/Qualifiers			
FT	Peptide	1..23			
FT		/label= leader peptide			
FT	Region	24..117			
FT		/label= variable-like region 1			
FT	Region	118..134			
FT		/label= joining-like region 1			
FT	Region	135..189			
FT		/label= variable-like region 2			
FT	Region	190..205			
FT		/label= joining-like region 2			
FT	Region	206..287			
FT		/label= variable-like region 3			
FT	Region	288..309			
FT		/label= joining-like region 3			
FT	Modified-site	297..299			
FT		/label= putative N-glycosylation site			
FT	Region	310..337			
FT		/label= variable-like region 4			
FT	Modified-site	325..327			
FT		/label= putative N-glycosylation site			
FT	Region	378..397			
FT		/label= joining-like region 4			
FT	Region	398..420			
FT		/label= trans-membrane region			
FT	Region	421..458			
XX		/label= cytoplasmic region			
FN	W08801304-A.				
XX	25-FEB-1988.				

XX 20-AUG-1987; 87WO-US002050.  
 PF  
 XX  
 PR 21-AUG-1986; 86US-00898587.  
 XX  
 PA (UYCO-) COLUMBIA UNIV.  
 XX (MADD/) MADDON P J.  
 XX  
 PI Litman DR, Maddon PJ, Chase L, Axel R, Weiss R, McDougal JS;  
 XX MPI. 1988-064019/09.  
 DR N-PSDB; AAN80512.  
 XX  
 PT Nucleic acid encoding T4 glyco:protein - used for treatment of AIDS and  
 PT producing antibodies for use as vaccine for immunisation against AIDS.  
 XX  
 PS Disclosure; Page 7; 128pp; English.  
 CC  
 CC T4 protein encoded by part of 3kb insert from human T cell library  
 CC (pt4B). (Updated on 25-MAR-2003 to correct PA field.)  
 XX  
 SQ Sequence 458 AA;

Query Match 87.7%; Score 2038; DB 1; Length 458;  
 Best Local Similarity 89.4%; Pred. No. 6e-131;  
 Matches 406; Conservative 2; Mismatches 16; Indels 30; Gaps 2;

QY 1 MNRGVPFRHLILVQLALLPAATQGNKVVLGKKGDTVELCTASOKKSIOFHMKNNSQIK 60  
 DB 1 MNRGVPFRHLILVQLALLPAATQGNKVVLGKKGDTVELCTASOKKSIOFHMKNNSQIK 60  
 QY 61 ILNGSGFLTKGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEVOL 120  
 DB 61 ILNGSGFLTKGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEVOL 120  
 QY 121 LVFGITANSPTHLLOGSITLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 180  
 DB 121 LVFGITANSPTHLLOGSITLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 180  
 QY 181 TWTCVTLNOKKVEFKIDIVLAFQKASSIYKKEGEVPSFPLAFVETKLTSGELMW 240  
 DB 181 TWTCVTLNOKKVEFKIDIVLAFQKASSIYKKEGEVPSFPLAFVETKLTSGELMW 240  
 QY 241 QAERASSSKSWITPDLKNKEVSVKRVTPDKLQMGKULPLHLTLFQALPOYAGSGLTLA 300  
 DB 241 QAERASSSKSWITPDLKNKEVSVKRVTPDKLQMGKULPLHLTLFQALPOYAGSGLTLA 300  
 QY 301 LEAKTGKLGHOEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKENKAKVSKREKPVW 360  
 DB 301 LEAKTGKLGHOEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKENKAKVSKREKAVW 360  
 QY 361 LNPEAGMMQCLSDSGOVLLESINIKVLPWTSPVPRASALPAPPTGSALDPQASALP 420  
 DB 361 LNPEAGMMQCLSDSGOVLLESINIKVLPWTSPVPRASALPAPPTGSALDPQASALP 420  
 QY 421 DPPASALPALAVISFLIGLIGV-ACVLARTR 453  
 DB 421 -----MALIVLGVAGLLFLTGLGIFCVRCRHR 425

RESULT 20  
 AAP91369  
 ID AAP91369 standard; protein; 458 AA.

XX AAP91369;  
 AC  
 XX 24-OCT-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 08-JAN-1990 (first entry)  
 XX  
 DE T4 protein.  
 XX  
 KW T4 protein; human immunodeficiency virus; AIDS; clone pt4B.

XX Homo sapiens; (human).  
 OS  
 XX  
 FH Key  
 FT Domain 1..23  
 FT Domain 24..117  
 FT Region 41  
 FT Region 109  
 FT Domain 118..134  
 FT Domain 135..189  
 FT Region 155  
 FT Region 184  
 FT Domain 190..204  
 FT Domain 205..286  
 FT Domain 287..309  
 FT Binding-site 296..298  
 FT Binding-site 310..376  
 FT Binding-site 325..327  
 FT Region 328  
 FT Region 370  
 FT Domain 377..397  
 FT Domain 398..420  
 FT Domain 421..458

EP330227-A.  
 30-AUG-1989.

24-FEB-1989; 89EP-00103297.

24-FEB-1988; 88US-00160348.

(UYCO-) COLUMBIA UNIV.  
 (SMIK ) SMITHKLINE BECKMAN CORP.  
 (UYCO ) UNIV COLUMBIA NEW YORK.  
 (UYCO ) UNIV COLUMBIA NEW YORK.

Maddon PJ, Axel R, Sweet RW, Arthos J;

WPI; 1989-250337/35.

Soluble T4 polypeptide derive. - inhibitors of extracellular and cell to cell spread of HIV used in prevention and treatment of AIDS.

Claim 1; Fig 6; 73pp; English.

T4 protein (AAN0619) inhibits extracellular and cell-to-cell spread of HIV. The therapeutic agent consists of amino acids +3-+185 fused to +351-+369; +3-+106 fused to +351-+369; or +3-+185. Also used to identify inhibitors of T4+ interactions, as target carrier proteins, and to generate monoclonal antibodies. Above features are: Domain 1 (D); starting at the N-terminal; D5 = V2; D5 = V3; D6 = V3; D7 = V4; D8 = V4; D9 = transmembrane; D10 = cytoplasmic; Regions are extracellular cysteines; and the 2 sites are potential N-linked glycosylation sites. (Updated on 25-MAR-2003 to correct PA field.) (Updated on 24-OCT-2003 to standardise OS field)

Sequence 458 AA;

Query Match 87.7%; Score 2038; DB 1; Length 458;  
 Best Local Similarity 89.4%; Pred. No. 6e-131;  
 Matches 406; Conservative 2; Mismatches 16; Indels 30; Gaps 2;

QY 1 MNRGVPFRHLILVQLALLPAATQGNKVVLGKKGDTVELCTASOKKSIOFHMKNNSQIK 60  
 DB 1 MNRGVPFRHLILVQLALLPAATQGNKVVLGKKGDTVELCTASOKKSIOFHMKNNSQIK 60  
 QY 61 ILNGSGFLTKGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEVOL 120  
 DB 61 ILNGSGFLTKGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEVOL 120  
 QY 121 LVFGITANSPTHLLOGSITLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 180

```

Db      121 LVFGLTANSDBTHLLOGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Qy      181 |TWTCTVLONOKKVEFKIDIVVLAFOKASSIVYKKEGEQVEFPLAFTVEKLTSGGELMW 240
Db      181 |TWTCTVLONOKKVEFKIDIVVLAFOKASSIVYKKEGEQVEFPLAFTVEKLTSGGELMW 240
Qy      241 QAERASSSKSWITFDLKNKEVSVKRVTDQPKLQMGKKLPLHLTLFQALPQYAGSGNLTIA 300
Db      241 QAERASSSKSWITFDLKNKEVSVKRVTDQPKLQMGKKLPLHLTLFQALPQYAGSGNLTIA 300
Qy      301 LEAKTGKLGHOEVNLYVMRATQLOKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPVVW 360
Db      301 LEAKTGKLGHOEVNLYVMRATQLOKNLTCEVWGPTSPKMLSLKENKEAKVSKREKAVVW 360
Qy      361 LNPEAGMOCCLSDSGOVLLESNIKLPPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
Db      361 LNPEAGMOCCLSDSGOVLLESNIKLPPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
Qy      421 DPPASALPALAVISFLGLGLGV-ACVLARTR 453
Db      397 -----MALIVLGAVAGLLFLGLGIFFCVRCRHR 425

```

RESULT 21  
AAY39826  
ID AAY39826 standard; protein; 458 AA.

```

AC      AAY39826;
XX
DT      03-DEC-1999 (first entry)
XX
DE      Soluble human T4 protein.
XX
KW      Soluble T4 protein; sT4; human; HIV; binding inhibitor; T4+ cell; AIDS;
KM      vaccine; immunisation; therapy.
OS      Homo sapiens.
XX
PN      US5958678-A.
XX
PD      28-SEP-1999.
XX
PF      12-DEC-1994; 94US-00354452.
XX
PR      21-AUG-1986; 86US-00898587.
PR      11-JUN-1991; 91US-00713564.
PR      06-JUL-1992; 92US-00909021.
XX
PA      (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI      McDougal JS, Weiss R, Axel R, Littman DR, Maddon PJ, Chess L;
XX      WPI; 1999-561025/47.
DR      N-PSDB; AA220695.
XX
PT      Human T4 protein inhibits HIV binding to T4 cells, useful for treating
XX      AIDS.
XX
PS      Example 3; Fig 6; 58pp; English.
XX
CC      This sequence represents the soluble human T4 protein of the invention.
CC      The soluble human T4 protein blocks the binding of HIV to T4+ cells and
CC      is therefore useful for the treatment of AIDS. Monoclonal antibodies
CC      against the T4 protein may be used as vaccines for immunising subjects
CC      against AIDS
XX
SQ      Sequence 458 AA;

```

Query Match 87.7%; Score 2038; DB 2; Length 458;  
Best Local Similarity 89.4%; Pred. No. 6e-131;  
Matches 406; Conservative 2; Mismatches 16; Indels 30; Gaps 2;  
Qy 1 NMRGVFPRHLVLVQLALPLPATQGNKVVLGKKGDTVELTCTASQKSIQFPMKNSNQIK 60

```

Db      1 |NMRGVFPRHLVLVQLALPLPATQGNKVVLGKKGDTVELTCTASQKSIQFPMKNSNQIK 60
Qy      61 ILGNQSFLLTKGPSKLNDRADSRRLMDQGNFPLIINKLKIEDSTYICEVEDQKEEYOL 120
Db      61 ILGNQSFLLTKGPSKLNDRADSRRLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEYOL 120
Qy      121 LVFGLTANSDBTHLLOGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Db      121 LVFGLTANSDBTHLLOGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Qy      181 |TWTCTVLONOKKVEFKIDIVVLAFOKASSIVYKKEGEQVEFPLAFTVEKLTSGGELMW 240
Db      181 |TWTCTVLONOKKVEFKIDIVVLAFOKASSIVYKKEGEQVEFPLAFTVEKLTSGGELMW 240
Qy      241 QAERASSSKSWITFDLKNKEVSVKRVTDQPKLQMGKKLPLHLTLFQALPQYAGSGNLTIA 300
Db      241 QAERASSSKSWITFDLKNKEVSVKRVTDQPKLQMGKKLPLHLTLFQALPQYAGSGNLTIA 300
Qy      301 LEAKTGKLGHOEVNLYVMRATQLOKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPVVW 360
Db      301 LEAKTGKLGHOEVNLYVMRATQLOKNLTCEVWGPTSPKMLSLKENKEAKVSKREKAVVW 360
Qy      361 LNPEAGMOCCLSDSGOVLLESNIKLPPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
Db      361 LNPEAGMOCCLSDSGOVLLESNIKLPPTWSTPVHPRASALPAPPTGSALPDPQTASALP 420
Qy      421 DPPASALPALAVISFLGLGLGV-ACVLARTR 453
Db      397 -----MALIVLGAVAGLLFLGLGIFFCVRCRHR 425

```

RESULT 22  
AAB19510  
ID AAB19510 standard; protein; 481 AA.

```

AC      AAB19510;
XX
DT      09-JAN-2001 (first entry)
XX
DE      CD4-IgM fusion protein CH4Pmu.
XX
KW      CD4; IgM; human; CD4Pmu; fusion protein; immunoglobulin; HIV; sIV; gp120;
KM      therapy; diagnosis.
OS      Homo sapiens.
XX
PN      US6117656-A.
XX
PD      12-SEP-2000.
XX
PF      07-JUN-1995; 95US-00479353.
XX
PR      22-JAN-1988; 88US-00147351.
PR      23-JAN-1989; 89US-00299596.
PR      09-JUN-1992; 92US-00886781.
PR      12-APR-1993; 93US-00057952.
PR      04-FEB-1994; 94US-00191708.
XX
PA      (GEHO ) GEN HOSPITAL CORP.
XX
PI      Seed B;
XX
DR      WPI; 2000-586558/55.
DR      N-PSDB; AAA50662.
XX
PT      CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or

```





```

AAR89450
ID AAR89450 standard; peptide; 398 AA.
XX
AC AAR89450;
XX
DT 26-SEP-1996 (first entry)
XX
DE CD4 D1-D4 domains.
XX
KM CD7; transmembrane domain; chimeric receptor; CD5; CD34; CH2; CH3; IgG1;
KW human; CD4; HIV; proteinaceous alpha-helix; T cell; B cell; neutrophil;
KW dendritic cell; therapy; mammal; infection.
XX
OS Homo sapiens.
XX
FN WO9603863-A1.
XX
PD 15-FEB-1996.
XX
PF 26-JUL-1995; 95MO-US009468.
XX
PR 02-AUG-1994; 94US-00284391.
XX
PR 24-FEB-1995; 95US-00394388.
XX
PA (GEMO ) GEN HOSPITAL CORP.
XX
PI Seed B, Banapour B, Romeo C, Kolanus W;
XX
DR WPI; 1996-129034/13.
XX
DR N-PSDB; AAT10797.
XX
PT Membrane-bound chimeric receptor comprising extracellular portion
XX including CD4 fragment - cells expressing receptor can be used for
XX treatment of HIV infection.
XX
Example 10; Fig 23; 134pp; English.
XX
PS This sequence represents the D1-D4 domains of CD4. This sequence is
XX included in the membrane bound proteinaceous chimeric receptor of the
XX invention. The extracellular portion of the chimeric receptor contains a
XX fragment of CD4 (amino acids 1-394 or 1-200 of the CD4 sequence) which
XX specifically recognises and binds HIV-infected cells, but does not
XX mediate HIV infection. The extracellular domain of the receptor is
XX separated from the cell membrane by 48 or 72 angstroms, or by one or more
XX proteinaceous alpha-helices. The transmembrane region of the chimeric
XX receptor contains a portion of the CD7, CD5 or CD34 transmembrane domain.
XX Alternatively, the extracellular portion of the receptor can also be
XX separated from the intracellular domain by the hinge, CH2 and CH3 domains
XX of human IgG1. The cells expressing the receptor are preferably T cells,
XX B cells, neutrophils, or dendritic cells. The therapeutic cells
XX expressing the chimeric receptor are administered to a mammal to treat
XX HIV infection
XX
SQ Sequence 398 AA;
XX
Query Match 87.6%; Score 2037; DB 2; Length 398;
XX Best Local Similarity 100.0%; Pred. No. 5.9e-13;
XX Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
1 NMRGVFRLHLVLQALPAATQGNKVVLGGKDPVELTCTASOKKSIQFMKNSNOIK 60
1 NMRGVFRLHLVLQALPAATQGNKVVLGGKDPVELTCTASOKKSIQFMKNSNOIK 60
1 NMRGVFRLHLVLQALPAATQGNKVVLGGKDPVELTCTASOKKSIQFMKNSNOIK 60
1 IINGNQSFLTKGPSKLNDRADSRSLMDQGNFPLIIKNKIBSDPTIYICEVDQKEEYVL 120
61 IINGNQSFLTKGPSKLNDRADSRSLMDQGNFPLIIKNKIBSDPTIYICEVDQKEEYVL 120
61 IINGNQSFLTKGPSKLNDRADSRSLMDQGNFPLIIKNKIBSDPTIYICEVDQKEEYVL 120
61 IINGNQSFLTKGPSKLNDRADSRSLMDQGNFPLIIKNKIBSDPTIYICEVDQKEEYVL 120
121 LVFGILTANSDTHLQOQSITLTLESPPGSSPSVQCRSPGKXIOGKTLISVQLBLQDSG 180
121 LVFGILTANSDTHLQOQSITLTLESPPGSSPSVQCRSPGKXIOGKTLISVQLBLQDSG 180
121 LVFGILTANSDTHLQOQSITLTLESPPGSSPSVQCRSPGKXIOGKTLISVQLBLQDSG 180
121 LVFGILTANSDTHLQOQSITLTLESPPGSSPSVQCRSPGKXIOGKTLISVQLBLQDSG 180
181 TWTCTVLQOQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFTVEXLTGSGELMW 240
181 TWTCTVLQOQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFTVEXLTGSGELMW 240
181 TWTCTVLQOQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFTVEXLTGSGELMW 240

```

```

DB 181 TWTCTVLQOQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPPLAFTVEXLTGSGELMW 240
QY 241 QAERASSSKSWTTPDLKKEVSVKRVTODPKIQMGKKLPPLHTTLPQALPQVAGSNLTLA 300
DB 241 QAERASSSKSWTTPDLKKEVSVKRVTODPKIQMGKKLPPLHTTLPQALPQVAGSNLTLA 300
QY 301 LEAKTGKLEHGVNLVVMRATOLQKXLTCEWGPSPKMLSLKLENKEAKSKREKPVWV 360
DB 301 LEAKTGKLEHGVNLVVMRATOLQKXLTCEWGPSPKMLSLKLENKEAKSKREKPVWV 360
QY 361 LNPEAGMMQCLSDSGVLLBSNINVLPTWSTPVH 395
DB 361 LNPEAGMMQCLSDSGVLLBSNINVLPTWSTPVH 395
DB 361 LNPEAGMMQCLSDSGVLLBSNINVLPTWSTPVH 395
RESULT 25
AAR27278
ID AAR27278 standard; protein; 532 AA.
XX
AC AAR27278;
XX
DT 25-MAR-2003 (revised)
XX
DT 28-JUL-1995 (first entry)
XX
DE CD4:gamma peptide chimeric protein.
XX
XX Fusion protein; CD4; extracellular domain; zeta; eta; gamma;
XX membrane spanning domain; intracellular domain; type I;
XX integral membrane homodimer; TCR; T cell antigen receptor;
XX extracellular domain; mouse; human; receptor; chimera;
XX Hsp-ALU tumour cell line; natural killer cell.
XX
OS Homo sapiens.
XX
FN WO9215322-A1.
XX
PD 17-SEP-1992.
XX
PF 06-MAR-1992; 92MO-US001785.
XX
PR 07-MAR-1991; 91US-00665961.
XX
PA (GEMO ) GEN HOSPITAL CORP.
XX
PI Seed B, Romeo C, Kolanus W;
XX
DR WPI; 1992-331474/40.
XX
DR N-PSDB; AAQ28706.
XX
PT Therapeutic cells expressing chimeric receptors - directing cellular
XX response to an infective agent, useful in treating HIV-1, AIDS
XX Pneumocystis carinii infections etc.
XX
PS Example 2; Page 74-76; 114pp; English.
XX
XX This sequence represents a fusion protein between the CD4 extracellular
XX domain and the gamma protein membrane spanning domain and intracellular
XX domain. The Fc-receptor-associated gamma chain is expressed in cell
XX surface complexes with additional polypeptides, some of which mediate
XX ligand recognition, and others which have undefined function. Gamma bears
XX a homodimeric structure and overall organisation very similar to that of
XX zeta (see also AAQ28704), and is a component of both the mast
XX cell/basophil high affinity IgE receptor, Fc-epsilon-RI, which consists
XX of at least three distinct polypeptide chains and one of the low affinity
XX receptors for IgG, represented in mice by Fc-gamma-RII-alpha. In the
XX production of the CD4 receptor chimera, the gamma cDNA was isolated from
XX the Hsp-ALU tumour cell line and from human natural killer cells. The
XX gamma cDNA was joined to the extracellular domain by engineering a BamHI
XX site just upstream of the membrane spanning domain, by a BamHI site
XX naturally present a few residues upstream of the membrane spanning
XX domain. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 532 AA;

```

Query Match 87.6%; Score 2037; DB 2; Length 532;  
 Best Local Similarity 100.0%; Pred. No. 8,4e-131;  
 Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 MNRGVPRHLLLVQLALLPAATQGNKVILGKGGDTVELTCTASQKSIQFHMKNNOIK 60
Db 1 MNRGVPRHLLLVQLALLPAATQGNKVILGKGGDTVELTCTASQKSIQFHMKNNOIK 60
Oy 61 ILNGSGFLTKGSPSKLNDRADSRSLMDQGNPILIKNLKIEDSDTYICEVEDQKEEVOL 120
Db 61 ILNGSGFLTKGSPSKLNDRADSRSLMDQGNPILIKNLKIEDSDTYICEVEDQKEEVOL 120
Oy 121 LVFGLTANSDTHLLQGGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Db 121 LVFGLTANSDTHLLQGGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Oy 181 TWTCVTIQQNKKEFKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGSGELMW 240
Db 181 TWTCVTIQQNKKEFKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGSGELMW 240
Oy 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOLPYAGSGNLTLA 300
Db 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOLPYAGSGNLTLA 300
Oy 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPVW 360
Db 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPVW 360
Oy 361 LNPEAGMOCCLSDSGVLLSNIKVLPTWSTPVH 395
Db 361 LNPEAGMOCCLSDSGVLLSNIKVLPTWSTPVH 395

```

RESULT 26  
 AAR78678  
 ID AAR78678 standard; protein: 532 AA.

```

XX AC AAR78678;
XX DT 16-APR-1996 (first entry)
XX DE T-cell receptor eta.
XX KW Chimeric receptor; CD4; T-cell receptor eta; HIV; cytolysis;
XX KW human immunodeficiency virus; adoptive immunotherapy.
XX OS Homo sapiens.
XX PN WO9521528-A1.
XX PD 17-AUG-1995.
XX PF 12-JAN-1995; 95WO-US000454.
XX PR 14-FEB-1994; 94US-00195395.
XX PR 02-AUG-1994; 94US-00284391.
XX PA (GCHO) GEN HOSPITAL CORP.
XX PI Seed B, Banapour B, Romeo C, Kojanus W;
XX DR WPI; 1995-292893/38.
XX DR N-PSDB; AAQ96124.
XX PT Target cytolysis of HIV-infected cells - by chimeric CD4 receptor-bearing
XX PT cells.
XX PS Example 2; Page 78-79; 118pp; English.
XX CC Fusion proteins comprising the extracellular domain of CD4 fused to T-
XX CC cell receptor zeta, gamma or eta (AAR78676-78, respectively) were
XX CC expressed in CV1 using a vaccine virus vector. These CD4:zeta, CD4:gamma

```

CC and CD4:eta chimeric receptors mediated cytolysis of targets expressing  
 CC HIV gp120/41  
 XX Sequence 532 AA;

Query Match 87.6%; Score 2037; DB 2; Length 532;  
 Best Local Similarity 100.0%; Pred. No. 8,4e-131;  
 Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 MNRGVPRHLLLVQLALLPAATQGNKVILGKGGDTVELTCTASQKSIQFHMKNNOIK 60
Db 1 MNRGVPRHLLLVQLALLPAATQGNKVILGKGGDTVELTCTASQKSIQFHMKNNOIK 60
Oy 61 ILNGSGFLTKGSPSKLNDRADSRSLMDQGNPILIKNLKIEDSDTYICEVEDQKEEVOL 120
Db 61 ILNGSGFLTKGSPSKLNDRADSRSLMDQGNPILIKNLKIEDSDTYICEVEDQKEEVOL 120
Oy 121 LVFGLTANSDTHLLQGGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Db 121 LVFGLTANSDTHLLQGGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Oy 181 TWTCVTIQQNKKEFKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGSGELMW 240
Db 181 TWTCVTIQQNKKEFKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGSGELMW 240
Oy 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOLPYAGSGNLTLA 300
Db 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOLPYAGSGNLTLA 300
Oy 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPVW 360
Db 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPVW 360
Oy 361 LNPEAGMOCCLSDSGVLLSNIKVLPTWSTPVH 395
Db 361 LNPEAGMOCCLSDSGVLLSNIKVLPTWSTPVH 395

```

RESULT 27  
 AAR89458  
 ID AAR89458 standard; protein: 532 AA.

```

XX AC AAR89458;
XX DT 26-SEP-1996 (first entry)
XX DE CD4:eta fusion protein.
XX KW CD7; transmembrane domain; chimeric receptor; CD5; CD34; CH2; CH3; IgG1;
XX KW human; CD4; HIV; proteinaceous alpha-helix; T cell; B cell; neutrophil;
XX KW dendritic cell; therapy; mammal; infection.
XX OS Synthetic.
XX PN WO9603883-A1.
XX PD 15-FEB-1996.
XX PF 26-JUL-1995; 95WO-US009468.
XX PR 02-AUG-1994; 94US-00284391.
XX PR 24-FEB-1995; 95US-00394388.
XX PA (GCHO) GEN HOSPITAL CORP.
XX PI Seed B, Banapour B, Romeo C, Kojanus W;
XX DR WPI; 1996-129034/13.
XX DR N-PSDB; AAT10803.
XX PT Membrane-bound chimeric receptor comprising extracellular portion
XX PT including CD4 fragment - cells expressing receptor can be used for
XX PT treatment of HIV infection.

```

XX Example 2; Page 80-81; 134pp; English.  
 PS  
 CC AAT10801-110803 represent membrane bound proteinaceous chimeric receptors  
 CC of the invention. This sequence represents the CD4:eta chimera. The  
 CC transmembrane region of the chimeric receptor acts to separate the  
 CC intracellular and extracellular domains of the chimera, and contains a  
 CC portion of the CD7 (see AAR89440), CD5 or CD34 transmembrane domain.  
 CC Alternatively, the extracellular portion of the receptor can be separated  
 CC from the intracellular domain by the hinge, CH2 and CH3 domains of human  
 CC IgG1 (see AAR89441). The extracellular portion of the chimeric receptor  
 CC contains a fragment of CD4 (amino acids 1-394 or 1-200 of the CD4  
 CC sequence, see AAR89450 and AAR89451) which specifically recognises and  
 CC binds HIV-infected cells, but does not mediate HIV infection. The  
 CC extracellular domain of the receptor is separated from the cell membrane  
 CC by 48 or 72 angstroms, or by one or more proteinaceous alpha-helices. The  
 CC cells expressing the receptor are preferably T cells, B cells,  
 CC neutrophils, or dendritic cells. The therapeutic cells expressing the  
 CC chimeric receptor are administered to a mammal to treat HIV infection  
 CC  
 SQ Sequence 532 AA;

Query Match 87.6%; Score 2037; DB 2; Length 532;  
 Best Local Similarity 100.0%; Pred. No. 8.4e-131;  
 Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NMRGVPRHLLVLTALPAATQGNKVVLGKKGTVELTCTASQKSIQFHWKNSNQIK 60  
 DB 1 NMRGVPRHLLVLTALPAATQGNKVVLGKKGTVELTCTASQKSIQFHWKNSNQIK 60  
 QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIIKNIKIEDSDTYICEVEDQKEEVQL 120  
 DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIIKNIKIEDSDTYICEVEDQKEEVQL 120  
 QY 121 LVFGLTANSDFHLQGGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSDFHLQGGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 QY 121 LVFGLTANSDFHLQGGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSDFHLQGGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 QY 181 TWTCTVLTQNKQKVEFKIDIVLAFOKASSIVYKKEBQVEFSPFLAFTVEKLTGSGELMW 240  
 DB 181 TWTCTVLTQNKQKVEFKIDIVLAFOKASSIVYKKEBQVEFSPFLAFTVEKLTGSGELMW 240  
 QY 181 TWTCTVLTQNKQKVEFKIDIVLAFOKASSIVYKKEBQVEFSPFLAFTVEKLTGSGELMW 240  
 DB 181 TWTCTVLTQNKQKVEFKIDIVLAFOKASSIVYKKEBQVEFSPFLAFTVEKLTGSGELMW 240  
 QY 241 QAEBASSSKWITFDLKNKEVSVKRVTQDPKLOMGKCLPLHITLTPALPOYAGSGNLTLA 300  
 DB 241 QAEBASSSKWITFDLKNKEVSVKRVTQDPKLOMGKCLPLHITLTPALPOYAGSGNLTLA 300  
 QY 301 LEAKTGKLTQAEVNLVVRATQLOKNTLCEVWGPTSPKMLSLKLENKEAKVSKREKPVVW 360  
 DB 301 LEAKTGKLTQAEVNLVVRATQLOKNTLCEVWGPTSPKMLSLKLENKEAKVSKREKPVVW 360  
 QY 361 LNPEAGMOCCLSDSGQVLLLESNIKVLPTWSTPWH 395  
 DB 361 LNPEAGMOCCLSDSGQVLLLESNIKVLPTWSTPWH 395

RESULT 28

AAW02215  
 ID AAW02215 standard; protein; 532 AA.

AC AAW02215;

DT 16-OCT-2003 (revised)

DT 11-NOV-1996 (first entry)

XX CD4:T-cell receptor eta chain chimaeric receptor.

XX Chimaeric receptor: cellular immunity; adoptive immunotherapy; CD4;

XX human immunodeficiency virus type 1; HIV-1; AIDS; therapy;

XX T-cell receptor eta chain; cytotoxic T lymphocyte; CTL.

XX Homo; sapiens.

OS Mus sp.

OS Chimeric.

XX Key location/Qualifiers  
 FH Domain 1..393  
 FT /label= "Extracellular\_domain"  
 FT /note= "CD4 extracellular domain"  
 FT Region 394..396  
 FT /label= "linker"  
 FT /note= "encoding DNA contains a BamHI site used for  
 FT fusion construction"  
 FT Region 397..532  
 FT /note= "region of fusion derived from eta chain,  
 FT preferred signal-transducing portions for construction of  
 FT the invention are amino acids 421-532, 423-455, 438-455,  
 FT 461-494, 494-528 or 400-420"  
 FT Domain 400..437  
 FT /label= "Transmembrane domain"  
 FT /note= "eta chain transmembrane domain"  
 FT Domain 438..575  
 FT /label= "intracellular domain"  
 FT /note= "eta chain intracellular domain"

PN WO9625953-A1.

XX 29-AUG-1996.

XX 25-JAN-1996; 96WO-US001056.

XX 24-FEB-1995; 95US-00394176.

XX (GEHO ) GEN HOSPITAL CORP.

XX Seed B, Romeo C, Kolanus W;

XX WPI; 1996-402134/40.

XX N-PSDB; AAT36760.

XX Direction of cellular immune response using therapeutic cell expressing 2

PT chimaeric receptors - comprising region binding to target cell and region

PT that signals target cell destruction, or CD28 region, partic. for

PT eliminating HIV-infected cells.

PS Claim 7; Page 77-78; 120pp; English.

XX A chimaeric receptor (AAW02215) comprises the extracellular domain of an  
 CC engineered form of the CD4 cellular receptor for HIV and the  
 CC transmembrane and intracellular regions, including the cytolytic signal-  
 CC transducing portion, of the mouse T-cell receptor eta chain. It can be  
 CC obtd. by inserting a gene fusion (AAT36760) into a vaccinia virus vector  
 CC and expression in host cells. Chimaeric receptors comprising CD4 fused to  
 CC eta, eta (see also AAW02213) or Fc receptor gamma (see also AAW02214)  
 CC chains are capable of directing cytotoxic T lymphocytes to specifically  
 CC recognise and kill cells expressing HIV gp120, thus providing a therapy  
 CC for AIDS. (Updated on 16-OCT-2003 to standardise OS field)

SQ Sequence 532 AA;

Query Match 87.6%; Score 2037; DB 2; Length 532;  
 Best Local Similarity 100.0%; Pred. No. 8.4e-131;  
 Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NMRGVPRHLLVLTALPAATQGNKVVLGKKGTVELTCTASQKSIQFHWKNSNQIK 60  
 DB 1 NMRGVPRHLLVLTALPAATQGNKVVLGKKGTVELTCTASQKSIQFHWKNSNQIK 60  
 QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIIKNIKIEDSDTYICEVEDQKEEVQL 120  
 DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIIKNIKIEDSDTYICEVEDQKEEVQL 120  
 QY 121 LVFGLTANSDFHLQGGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSDFHLQGGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 QY 181 TWTCTVLTQNKQKVEFKIDIVLAFOKASSIVYKKEBQVEFSPFLAFTVEKLTGSGELMW 240

```

Db      181  |TWTCTVLQONOKKVEFKIDIVLAFQKASSIYKKEGQVFFSFLAFTVEKLTSGELMW 240
Qy      241  |QAERASSSSKSWITFDLKNKEVSVKRVTDPKLQWKKLPHLTLPOALPOYAGSGNLTLA 300
Db      241  |QAERASSSSKSWITFDLKNKEVSVKRVTDPKLQWKKLPHLTLPOALPOYAGSGNLTLA 300
Qy      301  |LEAKTGKLGHOEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPYWV 360
Db      301  |LEAKTGKLGHOEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPYWV 360
Qy      361  |LNPEAGMWQCLISDSGQVLLBSNLIKVLPTWSTPVH 395
Db      361  |LNPEAGMWQCLISDSGQVLLBSNLIKVLPTWSTPVH 395

RESULT 29
AAW83141 ID AAW83141 standard; protein; 532 AA.
AC      AAW83141;
XX      03-FEB-1999 (first entry)
DE      Chimeric receptor containing human eta polypeptide.
XX      Human; zeta; eta; gamma; membrane-bound chimeric receptor; infection;
XX      tumour; cancer cell; autoimmune-generated cell; T cell receptor; CD3;
XX      CD4; B cell receptor; Fc receptor; pathogen; bacterial; fungal;
XX      protozoan; viral.
XX      Synthetic.
OS      Homo sapiens.
XX      US5843728-A.
XX      01-DEC-1998.
XX      05-APR-1995; 95US-00417495.
XX      07-MAR-1991; 91US-00665961.
XX      06-MAR-1992; 92US-00847566.
XX      28-FEB-1994; 94US-00203866.
XX      (GEMO ) GEN HOSPITAL CORP.
XX      PA
XX      PI Romeo C, Kolanus W, Seed B;
XX      WI; 1999-044582/04.
XX      DR N-PSDB; AAV70157.
XX      PT Membrane-bound chimeric receptors - comprising extracellular portion
XX      of which recognises and binds a target cell and an intracellular portion of
XX      e.g. a T-cell receptor.
XX      PS Claim 11; Col 45-48; 57pp; English.
XX      CC The present invention describes DNA encoding a membrane-bound chimeric
XX      receptor comprising: (a) an extracellular portion that specifically
XX      recognises and binds a target cell or a target infective agent; and (b)
XX      an intracellular portion of a T-cell receptor CD3, zeta or eta
XX      polypeptide, a B-cell receptor polypeptide or an Fc receptor polypeptide.
XX      The present sequence represents a chimeric receptor containing the human
XX      eta polypeptide. Cells expressing chimeric receptors of the present
XX      invention can be administered to mammals in order to destroy pathogens
XX      (e.g. bacteria, fungi, protozoa or viruses, especially HIV), cancer cells
XX      or autoimmune-generated cells
XX      CC
XX      SQ Sequence 532 AA;
XX
Query Match      87.6%; Score 2037; DB 2; Length 532;
Best Local Similarity 100.0%; Pred. No. 8.4e-131;
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1  |MNRGVPFRHLLVQLALLPAATQGNKVLGKGDVLELTCTASQKKSIOFHMNSNOIK 60
Db      1  |MNRGVPFRHLLVQLALLPAATQGNKVLGKGDVLELTCTASQKKSIOFHMNSNOIK 60
Qy      61  |ILNGQGSFLLTKGPKLNDRADSRRLWDQGNFLLINLKIEDSDTYICEVEDQKEEVOL 120
Db      61  |ILNGQGSFLLTKGPKLNDRADSRRLWDQGNFLLINLKIEDSDTYICEVEDQKEEVOL 120
Qy      121  |LVFGLTNSDTHLLQGGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDG 180
Db      121  |LVFGLTNSDTHLLQGGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDG 180
Qy      181  |TWTCTVLQONOKKVEFKIDIVLAFQKASSIYKKEGQVFFSFLAFTVEKLTSGELMW 240
Db      181  |TWTCTVLQONOKKVEFKIDIVLAFQKASSIYKKEGQVFFSFLAFTVEKLTSGELMW 240
Qy      241  |QAERASSSSKSWITFDLKNKEVSVKRVTDPKLQWKKLPHLTLPOALPOYAGSGNLTLA 300
Db      241  |QAERASSSSKSWITFDLKNKEVSVKRVTDPKLQWKKLPHLTLPOALPOYAGSGNLTLA 300
Qy      301  |LEAKTGKLGHOEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPYWV 360
Db      301  |LEAKTGKLGHOEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPYWV 360
Qy      361  |LNPEAGMWQCLISDSGQVLLBSNLIKVLPTWSTPVH 395
Db      361  |LNPEAGMWQCLISDSGQVLLBSNLIKVLPTWSTPVH 395

RESULT 30
AAV51082 ID AAV51082 standard; protein; 616 AA.
AC      AAV51082;
XX      23-MAR-2000 (first entry)
DE      Human fusion protein CD4Bgamma1.
XX      Human protein; human; CD4; IgG1; immunoglobulin; gp120;
XX      anti-human immunodeficiency virus; CD4Bgamma1.
XX      Homo sapiens.
XX      Synthetic.
XX      US6004781-A.
XX      21-DEC-1999.
XX      04-FEB-1994; 94US-00191708.
XX      22-JAN-1988; 88US-00147351.
XX      23-JAN-1989; 89US-00299596.
XX      09-JUN-1992; 92US-00896781.
XX      12-APR-1993; 93US-00057952.
XX      (GEMO ) GEN HOSPITAL CORP.
XX      PA
XX      PI Seed B;
XX      WI; 2000-085792/07.
XX      DR N-PSDB; AAZ44065.
XX      PT Fusion protein useful for the treatment of human immunodeficiency virus.
XX      PS Example 1; Col 59-70; 39pp; English.
XX      CC This invention describes a novel nucleic acid (I) encoding a fusion
XX      protein comprising a DNA sequence encoding amino acids 1-173 of CD4 (II)
XX      and a DNA sequence encoding a human immunoglobulin (Ig) heavy or light
XX      chain (III). The products of the invention have anti-human
XX      immunodeficiency virus (HIV) activity and are capable of binding to

```

CC gp120. The fusion protein is useful for treating human immunodeficiency  
 CC virus (HIV) or simian immunodeficiency virus (SIV). This sequence  
 CC represents the fusion protein CD4Bgamma1 which is constructed from CD4  
 CC linked to human IgG1 upstream of the hinge region  
 XX  
 SQ Sequence 616 AA;

Query Match 87.6%; Score 2037; DB 3; Length 616;  
 Best Local Similarity 96.6%; Pred. No. 1e-130;  
 Matches 399; Conservative 1; Mismatches 11; Indels 2; Gaps 1;

QY 1 MNRGVPRHLLVLTQALLPAATQGNKVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60  
 DB 1 MNRGVPRHLLVLTQALLPAATQGNKVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60  
 QY 61 ILNGSGSLTKGPKSLNDRADSRSLMDQGNFPLIKLKIEDSDTYICEVEDQKEEYOL 120  
 DB 61 ILNGSGSLTKGPKSLNDRADSRSLMDQGNFPLIKLKIEDSDTYICEVEDQKEEYOL 120  
 QY 121 LVFGLTANSDDHLLQGQSLTTLTSPGSSPSVOCSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSDDHLLQGQSLTTLTSPGSSPSVOCSPRGKNIQGGKTLVSQLELDQSG 180  
 QY 181 TWCTVLOQKQKVEFKIDIVLAFQKASSIYKKEGEQVEFSPLAFTVEKLTSGGELMW 240  
 DB 181 TWCTVLOQKQKVEFKIDIVLAFQKASSIYKKEGEQVEFSPLAFTVEKLTSGGELMW 240  
 QY 241 QAEKSSSSKSWITFDLKNKEVSVKRYTODPKLQMGKPLHLTLPOALPOYAGSGNLTLA 300  
 DB 241 QAEKSSSSKSWITFDLKNKEVSVKRYTODPKLQMGKPLHLTLPOALPOYAGSGNLTLA 300  
 QY 301 LEAKTGKLGHOEVNLYVMRATOLQKNTCEVWGPTSPKMLSLKLENKEAKYSKREKPYW 360  
 DB 301 LEAKTGKLGHOEVNLYVMRATOLQKNTCEVWGPTSPKMLSLKLENKEAKYSKREKPYW 360  
 QY 361 LNPEAGMOCCLSDSGQVLTLESNIKVLPTWSTPVH--PRASALPAPPTGSALP 411  
 DB 361 LNPEAGMOCCLSDSGQVLTLESNIKVLPTWSTPVHADPEAPBLGGPSVFLFP 413

## RESULT 31

AAVS9172  
 ID AAVS9172 standard; protein; 616 AA.  
 XX  
 AC AAVS9172;

XX 14-MAR-2000 (first entry)  
 DT  
 DE  
 XX CD4-Ig fusion protein CD4Bgamma1.  
 XX

KM HIV, extracellular; CD4; gp120; immunoglobulin; Ig; fusion protein;  
 KM secreted protein; SIV infection; medicament.  
 XX

OS Synthetic.  
 OS Homo sapiens.  
 XX

XX CA1340741-C.  
 PN

XX 14-SEP-1999.  
 PD

XX 20-JAN-1989; 89CA-00588749.  
 PF

XX 20-JAN-1989; 89CA-00588749.  
 PR

XX (GEHO ) GEN HOSPITAL CORP.  
 PA

XX Seed B;  
 PI

DR WPI; 2000-063015/06.  
 DR

XX N-PSDB; AA248205.  
 XX

PT New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in  
 the treatment of HIV or simian immunodeficiency virus infections.

XX Example 1; Page 61-68; 89pp; English.  
 XX

CC The invention provides a fusion gene encoding a fusion protein that  
 CC comprises an extracellular CD4 DNA sequence or its fragment which binds  
 CC to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA  
 CC sequence of an Ig heavy or light chain, where the DNA sequence encoding  
 CC the variable region has been replaced with the DNA sequence which encodes  
 CC extracellular CD4 or its gp120 binding fragment. The fusion protein is  
 CC capable of being secreted. The fusion proteins are useful for treating  
 CC HIV or SIV infections in animals, preferably humans. They are also useful  
 CC for producing medicaments which can be used for treating HIV or SIV  
 CC infections in humans. The present sequence represents the fusion protein  
 CC CD4Bgamma1 where the CD4 is linked to human IgG1 at the BamI site  
 CC downstream from the hinge region  
 XX

SQ Sequence 616 AA;

Query Match 87.6%; Score 2037; DB 3; Length 616;  
 Best Local Similarity 96.6%; Pred. No. 1e-130;  
 Matches 399; Conservative 1; Mismatches 11; Indels 2; Gaps 1;

QY 1 MNRGVPRHLLVLTQALLPAATQGNKVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60  
 DB 1 MNRGVPRHLLVLTQALLPAATQGNKVLGKGGDTVELTCTASQKSIQFHMKNNOIK 60  
 QY 61 ILNGSGSLTKGPKSLNDRADSRSLMDQGNFPLIKLKIEDSDTYICEVEDQKEEYOL 120  
 DB 61 ILNGSGSLTKGPKSLNDRADSRSLMDQGNFPLIKLKIEDSDTYICEVEDQKEEYOL 120  
 QY 121 LVFGLTANSDDHLLQGQSLTTLTSPGSSPSVOCSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSDDHLLQGQSLTTLTSPGSSPSVOCSPRGKNIQGGKTLVSQLELDQSG 180  
 QY 181 TWCTVLOQKQKVEFKIDIVLAFQKASSIYKKEGEQVEFSPLAFTVEKLTSGGELMW 240  
 DB 181 TWCTVLOQKQKVEFKIDIVLAFQKASSIYKKEGEQVEFSPLAFTVEKLTSGGELMW 240  
 QY 241 QAEKSSSSKSWITFDLKNKEVSVKRYTODPKLQMGKPLHLTLPOALPOYAGSGNLTLA 300  
 DB 241 QAEKSSSSKSWITFDLKNKEVSVKRYTODPKLQMGKPLHLTLPOALPOYAGSGNLTLA 300  
 QY 301 LEAKTGKLGHOEVNLYVMRATOLQKNTCEVWGPTSPKMLSLKLENKEAKYSKREKPYW 360  
 DB 301 LEAKTGKLGHOEVNLYVMRATOLQKNTCEVWGPTSPKMLSLKLENKEAKYSKREKPYW 360  
 QY 361 LNPEAGMOCCLSDSGQVLTLESNIKVLPTWSTPVH--PRASALPAPPTGSALP 411  
 DB 361 LNPEAGMOCCLSDSGQVLTLESNIKVLPTWSTPVHADPEAPBLGGPSVFLFP 413

## RESULT 32

AAAP93009  
 ID AAP93009 standard; protein; 631 AA.  
 XX

XX AAP93009;  
 AC

XX 25-MAR-2003 (revised)  
 DT

XX 02-NOV-1992 (first entry)  
 DE

XX Genetic construct which encodes CD4 linked to human IgG1 at the Bsp site  
 XX upstream of the hinge region (fusion protein CD4B-gamma-1).  
 DE

XX Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy;  
 XX diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.  
 KM

XX Homo sapiens.  
 OS

XX EP325262-A.  
 PN

XX 26-JUL-1989.  
 PD

XX 20-JAN-1989; 89EP-00100913.  
 PF

XX 22-JAN-1988; 88US-00147351.  
 XX (GEHO ) GEN HOSPITAL CORP.  
 XX Seed B;  
 XX WPI; 1989-214472/30.  
 XX N-PSDB; AAN90357.  
 XX Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV  
 XX infections or detecting HIV or SIV in sample.  
 XX Example; Table 2, Page 24-33; 68pp; English.  
 XX The fusion protein genes of the invention pref. comprise cDNA sequences  
 XX which encode CD4 or a fragment which binds gp120 ligated to an expression  
 XX plasmid which encodes an antibody in which the variable region of the  
 XX gene has been deleted (see WO87-02671). The CD4 portion of the fusion  
 XX protein may comprise the complete CD4 sequence, the 370 AA extracellular  
 XX region and the membrane spanning domain, or the extracellular region. The  
 XX Ig heavy chain is pref. from Igm, IgG1 or IgG3. The following are  
 XX specifically claimed: fusion proteins CD4-gamma-1, CD4mu, CD4mu, CD4E-  
 XX gamma1, and CD4mu (No. 67608), PCDA4-gamma (No. 67609) and PCDAE-gamma-1  
 XX (No. 67610). The plasmid containing (PCDAE-gamma-1) has been deposited in  
 XX E. coli (MC1061/P3) at the ATCC under accession number 67610. (Updated on  
 XX 25-MAR-2003 to correct PA field.)  
 XX Sequence 631 AA;

Query Match 87.5%; Score 2034.5; DB 1; Length 631;  
 Best Local Similarity 92.8%; Pred. No. 1.5e-130;  
 Matches 400; Conservative 1; Mismatches 21; Indels 9; Gaps 1;

QY 1 MNRGVPFRHLLVQLALPAAATGKRVLGKGGDTVELTCTASQKSIQPHMKNNOIK 60  
 DB 1 MNRGVPFRHLLVQLALPAAATGKRVLGKGGDTVELTCTASQKSIQPHMKNNOIK 60  
 QY ILNGGSLTKGSPSKLNDRAISRSLMDQGNPFLITIKLKIEDSPYICVEEDQKEEVOL 120  
 DB ILNGGSLTKGSPSKLNDRAISRSLMDQGNPFLITIKLKIEDSPYICVEEDQKEEVOL 120  
 QY 121 LVFGITANSPTHLLQGOSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSG 180  
 DB 121 LVFGITANSPTHLLQGOSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSG 180  
 QY 181 TWTCTVLOKQKVEKIDIVLAPQKASSIVYKKEGEVEFSFPLAFTVEKLTSGELMW 240  
 DB 181 TWTCTVLOKQKVEKIDIVLAPQKASSIVYKKEGEVEFSFPLAFTVEKLTSGELMW 240  
 QY 241 QAERASSKSMITFDLKKKEVSVKRVTPDKLQMGKPLHLTLPOALPOYAGSGNTLTA 300  
 DB 241 QAERASSKSMITFDLKKKEVSVKRVTPDKLQMGKPLHLTLPOALPOYAGSGNTLTA 300  
 QY 301 LEAKTGKLGHEVNLVVMRATQLOKXNLTCCEVWGPTSPKMLSLKLENKKAQVSKKEKPYV 360  
 DB 301 LEAKTGKLGHEVNLVVMRATQLOKXNLTCCEVWGPTSPKMLSLKLENKKAQVSKKEKPYV 360  
 QY 361 LNPEAGMOCCLSDSGOVLLESNTIKVLPWTSTPVH-----PRASALPAPRTGSLP 411  
 DB 361 LNPEAGMOCCLSDSGOVLLESNTIKVLPWTSTPVH-----PRASALPAPRTGSLP 411  
 QY 412 DPQTASALPDP 422  
 DB 421 GPSVFLFPKP 431

RESULT 33  
 ID AAB19508  
 XX AAB19508 standard; protein; 631 AA.  
 AC AAB19508;  
 XX

DT 09-JAN-2001 (first entry)  
 XX CD4-IgG1 fusion protein CH4Bgammal.  
 DE CD4; IgG1; human; CD4gamma1, fusion protein; immunoglobulin; HIV; SIV;  
 XX gp120; therapy; diagnosis.  
 KM Homo sapiens.  
 XX Key  
 FH Protein 1..395 Location/Qualifiers  
 FT /note="CD4 extracellular region"  
 FT 400..631  
 FT Protein /note="IgG1 heavy chain"  
 XX US6117656-A.  
 PN 12-SEP-2000.  
 XX 07-JUN-1995; 95US-00479353.  
 XX 22-JAN-1988; 88US-00147351.  
 PR 23-JAN-1989; 89US-00299596.  
 PR 09-JUN-1992; 92US-00896781.  
 PR 12-APR-1993; 93US-00057952.  
 PR 04-FEB-1994; 94US-00191708.  
 XX (GEHO ) GEN HOSPITAL CORP.  
 XX Seed B;  
 XX WPI; 2000-58658/55.  
 DR N-PSDB; AAS0661.  
 XX

PT CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or  
 PT SIV.  
 PS Example 1; Col 29-42; 39pp; English.

XX The present sequence is that of fusion protein CD4Bgammal comprising the  
 CC extracellular portion of CD4, which binds to HIV gp120, linked at its C-  
 CC terminus to the human IgG1 heavy chain. To obtain the fusion protein, DNA  
 CC encoding CD4 was linked to IgG1 DNA at the Bsp site upstream of the hinge  
 CC region (see AAS0661). Fusion protein CD4Bgammal and a nucleic acid  
 CC encoding it are claimed. Also claimed are a vector comprising the nucleic  
 CC acid, and a method of producing the fusion protein in secreted form using  
 CC a transformed host cell. The fusion protein may further comprise a  
 CC therapeutic agent, radiolabel or NMR imaging agent. The fusion protein  
 CC can be administered to an animal (including humans) for treatment of HIV  
 CC or SIV infection, and can also be used in assays for HIV or SIV, imaging  
 CC and tissue stains. IgG1 fusion proteins such as CD4Bgammal provide both  
 CC complement-mediated and cell-mediated immunity

XX Sequence 631 AA;

Query Match 87.5%; Score 2034.5; DB 3; Length 631;  
 Best Local Similarity 92.8%; Pred. No. 1.5e-130;  
 Matches 400; Conservative 1; Mismatches 21; Indels 9; Gaps 1;

QY 1 MNRGVPFRHLLVQLALPAAATGKRVLGKGGDTVELTCTASQKSIQPHMKNNOIK 60  
 DB 1 MNRGVPFRHLLVQLALPAAATGKRVLGKGGDTVELTCTASQKSIQPHMKNNOIK 60  
 QY ILNGGSLTKGSPSKLNDRAISRSLMDQGNPFLITIKLKIEDSPYICVEEDQKEEVOL 120  
 DB 61 ILNGGSLTKGSPSKLNDRAISRSLMDQGNPFLITIKLKIEDSPYICVEEDQKEEVOL 120  
 QY 121 LVFGITANSPTHLLQGOSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSG 180  
 DB 121 LVFGITANSPTHLLQGOSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSG 180  
 QY 181 TWTCTVLOKQKVEKIDIVLAPQKASSIVYKKEGEVEFSFPLAFTVEKLTSGELMW 240  
 DB 181 TWTCTVLOKQKVEKIDIVLAPQKASSIVYKKEGEVEFSFPLAFTVEKLTSGELMW 240

```

Db      181 TWCTVLQKQKVEFKIDIVLAFOKASSIYKKEGEVSEFPLAFTVEKLTGSGELMW 240
Qy      241 QAEKASSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGNLTLA 300
Db      241 QAEKASSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGNLTLA 300
Qy      301 LEAKTGKLGHOEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKLENKAKVSKREKPVWV 360
Db      301 LEAKTGKLGHOEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKLENKAKVSKREKPVWV 360
Qy      361 LNPEAGMMOCLSDSGOVLLESNIKVLPTWSTPPVH-----PRASALPAPPTGSALP 411
Db      361 LNPEAGMMOCLSDSGOVLLESNIKVLPTWSTPPVHADPEEPKSCDHTCPCPAPBELLG 420
Qy      412 DPQTASALPDP 422
Db      421 GPSVFLPPPKP 431

```

## RESULT 34

```

AA51079
ID      AA51079 standard; protein; 631 AA.
XX
AC      AA51079;
XX
DT      23-MAR-2000 (first entry)
XX
DE      Human fusion protein CD4Egammal.
XX
KW      Fusion protein; human; CD4; IgG1; immunoglobulin; gp120;
KW      anti-human immunodeficiency virus; CD4Egammal.
XX
OS      Homo sapiens.
XX
XX      Synthetic.
XX
XX      US6004781-A.
XX
XX      21-DEC-1999.
XX
PF      04-FEB-1994; 94US-00191708.
XX
PR      22-JAN-1988; 88US-00147351.
PR      23-JAN-1989; 89US-00299596.
PR      09-JUN-1992; 92US-00896781.
PR      12-APR-1993; 93US-00057952.
XX
XX      (GEHO ) GEN HOSPITAL CORP.
XX
XX      Seed B;
XX
XX      WPI: 2000-085792/07.
XX      N-PSDB; AA244062.
XX
XX      Fusion protein useful for the treatment of human immunodeficiency virus.
XX
XX      Example 1; Col 29-42; 39pp; English.
XX
XX      This invention describes a novel nucleic acid (I) encoding a fusion
XX      protein comprising a DNA sequence encoding amino acids 1-173 of CD4 (II)
XX      and a DNA sequence encoding a human immunoglobulin (Ig) heavy or light
XX      chain (III). The products of the invention have anti-human
XX      immunodeficiency virus (HIV) activity and are capable of binding to
XX      gp120. The fusion protein is useful for treating human immunodeficiency
XX      virus (HIV) or simian immunodeficiency virus (SIV). This sequence
XX      represents the fusion protein CD4Egammal which is constructed from CD4
XX      linked to human IgG1 upstream of the hinge region
XX
XX      Sequence 631 AA;

```

```

Query Match      87.5%; Score 2034.5; DB 3; Length 631;
Best Local Similarity 92.8%; Pred. No. 1,5e-130;
Matches 400; Conservative 1; Mismatches 21; Indels 9; Gaps 1;

```

```

Qy      1 NMRGVPFPHLLLVLTALLPAATQGNKVLGKKGDTVELTASQKSIQFMKNSNOIK 60
Db      1 NMRGVPFPHLLLVLTALLPAATQGNKVLGKKGDTVELTCTASQKSGIFMKNSNOIK 60
Qy      61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPLIINKLKI BSDPTYICEVEDQKEEYOL 120
Db      61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPLIINKLKI BSDPTYICEVEDQKEEYOL 120
Qy      121 LVFGLTANSDPHLLOGSLTTLTLESPPGSSPEVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Db      121 LVFGLTANSDPHLLOGSLTTLTLESPPGSSPEVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Qy      181 TWCTVLQKQKVEFKIDIVLAFOKASSIYKKEGEVSEFPLAFTVEKLTGSGELMW 240
Db      181 TWCTVLQKQKVEFKIDIVLAFOKASSIYKKEGEVSEFPLAFTVEKLTGSGELMW 240
Qy      241 QAEKASSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGNLTLA 300
Db      241 QAEKASSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGNLTLA 300
Qy      301 LEAKTGKLGHOEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKLENKAKVSKREKPVWV 360
Db      301 LEAKTGKLGHOEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKLENKAKVSKREKPVWV 360
Qy      361 LNPEAGMMOCLSDSGOVLLESNIKVLPTWSTPPVH-----PRASALPAPPTGSALP 411
Db      361 LNPEAGMMOCLSDSGOVLLESNIKVLPTWSTPPVHADPEEPKSCDHTCPCPAPBELLG 420
Qy      412 DPQTASALPDP 422
Db      421 GPSVFLPPPKP 431

```

## RESULT 35

```

AA59169
ID      AA59169 standard; protein; 631 AA.
XX
AC      AA59169;
XX
DT      14-MAR-2000 (first entry)
XX
DE      CD4-Ig fusion protein CD4Egammal.
XX
XX      HIV; extracellular; CD4; gp120; immunoglobulin; Ig; fusion protein;
XX      secreted protein; SIV infection; medicament.
XX
XX      Synthetic.
XX
XX      Homo sapiens.
XX
XX      CA1340741-C.
XX
XX      14-SEP-1999.
XX
XX      20-JAN-1989; 89CA-00588749.
XX
XX      20-JAN-1989; 89CA-00588749.
XX
XX      (GEHO ) GEN HOSPITAL CORP.
XX
XX      Seed B;
XX
XX      WPI: 2000-063015/06.
XX      N-PSDB; AA248202.
XX
XX      New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in
XX      the treatment of HIV or simian immunodeficiency virus infections.
XX
XX      Example 1; Page 37-46; 89pp; English.
XX
XX      The invention provides a fusion gene encoding a fusion protein that
XX      comprises an extracellular CD4 DNA sequence or its fragment which binds
XX      to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA
XX      sequence of an Ig heavy or light chain, where the DNA sequence encoding

```

CC the variable region has been replaced with the DNA sequence which encodes  
CC extracellular CD4 or its gp120 binding fragment. The fusion protein is  
CC capable of being secreted. The fusion proteins are useful for treating  
CC HIV or SIV infections in animals, preferably humans. They are also useful  
CC for producing medicaments which can be used for treating HIV or SIV  
CC infections in humans. The present sequence represents the fusion protein  
CC CD4gamma1 where the CD4 is linked to human IgG1 at the Esp site upstream  
CC of the hinge region  
XX  
XX

SQ Sequence 631 AA;

Query Match 87.5%; Score 2034.5; DB 3; Length 631;

Best Local Similarity 92.8%; Pred. No. 1.5e-130;

Matches 400; Conservative 1; Mismatches 21; Indels 9; Gaps 1;

```
QY 1 MNRGVPFRHLLVLTALPAATQGNKVLGKGGDTVELTCTASQKKSIOFHMKNNOIK 60
DB 1 MNRGVPFRHLLVLTALPAATQGNKVLGKGGDTVELTCTASQKKSIOFHMKNNOIK 60
QY 61 ILNGQSFLLTKGPKSLNDRADSRSLMDQGNFPLITKNLKIEDSDTYICEVEDQKEVQL 120
DB 61 ILNGQSFLLTKGPKSLNDRADSRSLMDQGNFPLITKNLKIEDSDTYICEVEDQKEVQL 120
QY 121 LVFGITANSPTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGITANSPTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 121 LVFGITANSPTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGITANSPTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 121 TWTCVLONOKKVEPKIDIVLAFQKASSIVYKKEGQVFSFPLAFTVEKLTGSGELMW 240
DB 121 TWTCVLONOKKVEPKIDIVLAFQKASSIVYKKEGQVFSFPLAFTVEKLTGSGELMW 240
QY 241 QAEKASSSSKSWITFDLKNKEVSVKRVTQDPKLOMGKPLHLTLPOALPOYAGSGNLTLA 300
DB 241 QAEKASSSSKSWITFDLKNKEVSVKRVTQDPKLOMGKPLHLTLPOALPOYAGSGNLTLA 300
QY 301 LEAKTGKLEHVEVNLVVMRATQLOKNLTCEWGPSPKLMSTLKENKAQVSKREKPVW 360
DB 301 LEAKTGKLEHVEVNLVVMRATQLOKNLTCEWGPSPKLMSTLKENKAQVSKREKPVW 360
QY 361 LNPBAGMWQCILSDSGVLLSENIKVLPTWSTPVH-----PRASALPAPPGSALP 411
DB 361 LNPBAGMWQCILSDSGVLLSENIKVLPTWSTPVHADPEEPKSCDKTTCPCPCPABELG 420
QY 412 DPGTASALPDP 422
DB 421 GPSVFLPPPKP 431
```

RESULT 36

AAR78673 standard; protein; 398 AA.

AAR78673;

12-APR-1996 (first entry)

CD4 domain DI-D4.

Chimeric receptor; CD4; T-cell receptor; HIV; cytolysis;

human immunodeficiency virus; adoptive immunotherapy.

Homo sapiens.

MO9521528-A1.

17-AUG-1995.

12-JAN-1995; 95WO-US000454.

14-FEB-1994; 94US-00195395.

02-AUG-1994; 94US-00284391.

(GEHO) GEN HOSPITAL CORP.

XX Seed B, Banapour B, Romeo C, Kolanus W;  
XX WPI: 1995-292893/38.  
DR N-PSDB; AAQ96103.  
XX Target cytolysis of HIV-infected cells - by chimeric CD4 receptor-bearing  
PT cells.  
XX  
XX

Example 10; Fig 23; 118pp; English.

CC Extracellular domains DI-D4 (AAR78673) or DI-D2 (AAR78674) of human CD4  
CC are used in the construction of chimeric receptors utilized in the  
CC targeted cytolysis of cells expressing HIV envelope proteins on their  
CC surface. The chimeric receptors comprise the extracellular domain (pref.  
CC amino acids 1-394 or 1-200) of CD4 linked to an intracellular portion,  
XX e.g. of T-cell receptor protein zeta  
XX

SQ Sequence 398 AA;

Query Match 87.5%; Score 2034; DB 2; Length 398;

Best Local Similarity 99.7%; Pred. No. 9.5e-131;

Matches 394; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MNRGVPFRHLLVLTALPAATQGNKVLGKGGDTVELTCTASQKKSIOFHMKNNOIK 60
DB 1 MNRGVPFRHLLVLTALPAATQGNKVLGKGGDTVELTCTASQKKSIOFHMKNNOIK 60
QY 61 ILNGQSFLLTKGPKSLNDRADSRSLMDQGNFPLITKNLKIEDSDTYICEVEDQKEVQL 120
DB 61 ILNGQSFLLTKGPKSLNDRADSRSLMDQGNFPLITKNLKIEDSDTYICEVEDQKEVQL 120
QY 61 ILNGQSFLLTKGPKSLNDRADSRSLMDQGNFPLITKNLKIEDSDTYICEVEDQKEVQL 120
DB 61 ILNGQSFLLTKGPKSLNDRADSRSLMDQGNFPLITKNLKIEDSDTYICEVEDQKEVQL 120
QY 121 LVFGITANSPTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGITANSPTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 121 LVFGITANSPTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGITANSPTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCVLONOKKVEPKIDIVLAFQKASSIVYKKEGQVFSFPLAFTVEKLTGSGELMW 240
DB 181 TWTCVLONOKKVEPKIDIVLAFQKASSIVYKKEGQVFSFPLAFTVEKLTGSGELMW 240
QY 241 QAEKASSSSKSWITFDLKNKEVSVKRVTQDPKLOMGKPLHLTLPOALPOYAGSGNLTLA 300
DB 241 QAEKASSSSKSWITFDLKNKEVSVKRVTQDPKLOMGKPLHLTLPOALPOYAGSGNLTLA 300
QY 301 LEAKTGKLEHVEVNLVVMRATQLOKNLTCEWGPSPKLMSTLKENKAQVSKREKPVW 360
DB 301 LEAKTGKLEHVEVNLVVMRATQLOKNLTCEWGPSPKLMSTLKENKAQVSKREKPVW 360
QY 361 LNPBAGMWQCILSDSGVLLSENIKVLPTWSTPVH 395
DB 361 LNPBAGMWQCILSDSGVLLSENIKVLPTWSTPVH 395
```

RESULT 37

AA88329 standard; protein; 458 AA.

AA88329;

14-JUN-2000 (first entry)

T4 glycoprotein amino acid sequence.

sT4; glycoprotein; human immunodeficiency virus; HIV; block binding;

AIDS; treatment; inhibit; cell to cell spread; infection; fusion.

Mammalia.

US5126433-A.

30-JUN-1992.

23-OCT-1987; 87US-00114244.



XX 21-AUG-1986; 86US-00898587.  
XX (UYCO ) UNIV COLUMBIA NEW YORK.  
XX  
XX Maddon PJ, Chess L, Axel R, Weiss R, Littman DR, McDougal JS;  
XX WPI; 2000-348913/30.  
XX Soluble T4 glycoprotein useful for prevention and treatment of acquired  
XX immunodeficiency syndrome and for screening inhibitors of human  
XX immunodeficiency viral binding.  
XX  
XX Example; Fig 6; 64pp; English.  
XX  
XX This sequence represents the amino acid sequence of glycosylated sT4  
XX glycoprotein. Human immunodeficiency virus (HIV) uses sT4 as a target  
XX receptor on T cells. The invention relates to glycosylated sT4 which  
XX functions by blocking the binding of HIV to T4 target cells, and can be  
XX used for the prophylaxis and treatment of AIDS patients. Administration  
XX of sT4 effectively inhibits the cell to cell spreading of HIV infection  
XX and also the fusion of HIV-infected T4 cells and non-infected T4 cells.  
XX The administration of T4 alleviates several symptoms associated with  
XX AIDS, and prevents the occurrence of new pathological changes. The sT4  
XX glycoprotein is useful for the prophylaxis and treatment of patients with  
XX AIDS. It is also useful as a reagent to identify natural, synthetic or  
XX recombinant molecules which act as therapeutic agents or inhibitors of  
XX T4+ cell interactions and in diagnostic assays for detection T4 proteins  
XX or molecules  
XX  
XX Sequence 458 AA;  
SQ

Query Match 87.4%; Score 2032; DB 3; Length 458;  
Best Local Similarity 89.2%; Pred. No. 1.5e-130; Indels 30; Gaps 2;  
Matches 405; Conservative 2; Mismatches 17;  
QY 1 NMRGVPFRHLVLVQLALLPAATQGNKVVLGKKGDTVELTCTASQKSIQFHMKNNOIX 60  
DB 1 NMRGVPFRHLVLVQLALLPAATQGNKVVLGKKGDTVELTCTASQKSIQFHMKNNOIX 60  
QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPPLIIKNLKIEDSDTYICEVEDQKEEYOL 120  
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPPLIIKNLKIEDSDTYICEVEDQKEEYOL 120  
QY 121 LVFGLTANSDFHLQGSQSLTLTLSPGSSPVQCSPPGKNIQGGKTLVSQLELDQSG 180  
DB 121 LVFGLTANSDFHLQGSQSLTLTLSPGSSPVQCSPPGKNIQGGKTLVSQLELDQSG 180  
QY 181 TWCTVLOQOKKEFIDIVLAFOKASSIVYKKEGOVEFSPLAFTVEKLTGSGELMW 240  
DB 181 TWCTVLOQOKKEFIDIVLAFOKASSIVYKKEGOVEFSPLAFTVEKLTGSGELMW 240  
QY 241 QABRASSSSSWITFDLKNKEVSVKRVTDPKLQMGKCLPLHLTLPOALPOYAGSGNLTIA 300  
DB 241 QABRASSSSSWITFDLKNKEVSVKRVTDPKLQMGKCLPLHLTLPOALPOYAGSGNLTIA 300  
QY 301 LEAKTCKLHOEVNLVVMRATLOLQKNTCEVMGPTSPKMLSLKLENKAKVSKREKPVV 360  
DB 301 LEAKTCKLHOEVNLVVMRATLOLQKNTCEVMGPTSPKMLSLKLENKAKVSKREKPVV 360  
QY 361 LNPEAGMOCLSDSGOVLLESNIKVLPTWSPVPHRASALPAPPTGSLPPQRTASALP 420  
DB 361 LNPEAGMOCLSDSGOVLLESNIKVLPTWSPVPHRASALPAPPTGSLPPQRTASALP 420  
QY 421 DEPPASALPALAVISFLLGLGIV-ACVLAART 453  
DB 397 -----MALIVLGGVAGLLFLFIGIGIFVCRCRR 425

RESULT 38  
AAB81502  
ID AAB81502 standard; protein; 458 AA.  
XX

AC AAB81502;  
XX  
XX 18-JUN-2001 (first entry)  
XX  
XX Human CD4 protein.  
XX  
XX Human; CD4; CD4 fusion protein; oligomerisation;  
XX receptor-ligand interaction inhibition; surface plasmon resonance; SPR;  
XX T cell receptor binding; MHC binding; carcinoma; autoimmune disease;  
XX multiple sclerosis; human immunodeficiency virus; HIV; diabetes;  
XX rheumatoid arthritis; immune disorder.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX FT Peptide 1..25  
XX FT /label= Signal\_peptide  
XX FT Protein 26..458  
XX FT /label= Human\_CD4  
XX  
XX WO200122084-A2.  
XX  
XX 29-MAR-2001.  
XX  
XX 18-SEP-2000; 2000MO-GB003579.  
XX  
XX 21-SEP-1999; 99GB-00022352.  
XX  
XX (AVID-) AVIDEX LTD.  
XX  
XX Jakobsen BK;  
XX  
XX WPI; 2001-273470/28.  
XX DB N-PSDB; AAF82582.  
XX  
XX Sequential screening of candidate compounds library for those which  
XX inhibit binding of low affinity receptor-ligand interaction having fast  
XX binding kinetics, using interfacial optical assay.  
XX  
XX Disclosure; Fig 13; 91pp; English.  
XX  
XX The present sequence is human CD4. Human CD4 extracellular domains 1 and  
XX 2 were used in the construction of CD4 oligomerisation fusion proteins.  
XX The fusion proteins contain an oligomerisation domain that enables the  
XX proteins to bind to one another to form oligomers. The oligomers may be  
XX used in an invention relating to a method for screening for compounds  
XX with the ability to inhibit a low affinity receptor-ligand interaction.  
XX The method uses an interfacial optical assay, such as surface plasmon  
XX resonance (SPR). The method is useful for screening candidate compounds  
XX for the ability to inhibit interaction between MHC/peptide complex and T  
XX cell receptor, and MHC/peptide complex and CD8 or CD4 co-receptor. The  
XX compounds identified by the above methods which interfere with T cell  
XX receptor binding to a particular HLA type molecule are useful as immune  
XX inhibitors for treating carcinomas, autoimmune diseases such as multiple  
XX sclerosis, human immunodeficiency virus (HIV) infection, rheumatoid  
XX arthritis, Hashimoto's disease, insulin dependent diabetes, Good  
XX pasture's syndrome, uveitis, psoriasis and graft rejection  
XX  
XX Sequence 458 AA;  
SQ

Query Match 87.4%; Score 2032; DB 4; Length 458;  
Best Local Similarity 89.2%; Pred. No. 1.5e-130; Indels 30; Gaps 2;  
Matches 405; Conservative 2; Mismatches 17;  
QY 1 NMRGVPFRHLVLVQLALLPAATQGNKVVLGKKGDTVELTCTASQKSIQFHMKNNOIX 60  
DB 1 NMRGVPFRHLVLVQLALLPAATQGNKVVLGKKGDTVELTCTASQKSIQFHMKNNOIX 60  
QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPPLIIKNLKIEDSDTYICEVEDQKEEYOL 120  
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPPLIIKNLKIEDSDTYICEVEDQKEEYOL 120  
QY 121 LVFGLTANSDFHLQGSQSLTLTLSPGSSPVQCSPPGKNIQGGKTLVSQLELDQSG 180

Db	121	LVFGLTNSDTHLLQGSGLTLTLESPPGSSPSVQCRSPGKNIQGKTLTSSQLELDSCG	180
Qy	181	TWTCTVLQNKKEVFKIDIVLVAFOKASSIVYKKEGEQVEFSFLAFTVEKLTSGELMW	240
Db	181	TWCTVLQNKKEVFKIDIVLVAFOKASSIVYKKEGEQVEFSFLAFTVEKLTSGELMW	240
Qy	241	QAEBASSKSWITFDLKNKEVSKRVTDPKLQMGKULPLHLTPQALPQYASGNTTLA	300
Db	241	QAEBASSKSWITFDLKNKEVSKRVTDPKLQMGKULPLHLTPQALPQYASGNTTLA	300
Qy	301	LEAKTGLTQHEVVLVYMRATQLOKNLTCEWGPSTPKMLSLKENNEAKYSKREPVWV	360
Db	301	LEAKTGLTQHEVVLVYMRATQLOKNLTCEWGPSTPKMLSLKENNEAKYSKREKAVWV	360
Qy	361	LNPEAGMOCGLSDSGOVLLESNIKVLPTWSTPVHPRASALPAPETGSAFLDPQTASALP	420
Db	361	LNPEAGMOCGLSDSGOVLLESNIKVLPTWSTPVQP-----	366
Qy	421	DPEASALPAAALAVISFLIGLGLGVACVLAATRR	453
Db	397	-----MALIVLGAVAGLLFTIGLGFVCRCHRR	425

XX	RESULT 39
XX	ADD25609
XX	ID ADD25609 standard; protein; 458 AA.
XX	AC
XX	ADD25609;
XX	DT
XX	15-JAN-2004 (first entry)
XX	DE Binding domain-immunoglobulin fusion protein-associated protein #92.
KW	Binding domain; immunoglobulin; fusion protein; cytosolic;
KW	antiarthritic; immunosuppressive; antidiabetic; anticholesteric;
KW	neuroprotective; hinge region; immunoglobulin heavy chain;
KW	CH2 constant region; CH3 constant region; 1961;
KW	antibody dependent cell-mediated cytotoxicity; ADCC; complement fixation;
KW	malignant condition; B-cell disorder; melanoma; carcinoma; sarcoma;
KW	rhematoid arthritis; myasthenia gravis; Grave's disease;
KW	type I diabetes mellitus; multiple sclerosis; autoimmune disease.
XX	Unidentified.
OS	
XX	US2003118592-A1.
PN	
XX	26-JUN-2003.
PD	
XX	25-JUL-2002; 2002US-00207655.
PF	
XX	17-JAN-2001; 2001US-0367358P.
PR	17-JAN-2002; 2002US-0005530.
PR	03-JUN-2002; 2002US-0385691P.
XX	(GENE-) GENE-CRAFT INC.
XX	PA
XX	Ledbetter JA, Hayden-Ledbetter MS, Thompson PA;
PI	
XX	WPI; 2003-801317/75.
DR	
XX	New binding domain-immunoglobulin fusion protein, useful for treating a
PT	subject having or suspected of having a malignant condition or a B-cell
PT	disorder, e.g. melanoma, Grave's disease or autoimmune disease.
XX	
PS	Disclosure; SEQ ID NO 170; 157bp; English.
XX	
CC	The invention relates to a binding domain-immunoglobulin fusion protein
CC	comprising a binding domain polypeptide that is fused to an
CC	immunoglobulin hinge region polypeptide, an immunoglobulin heavy chain
CC	CH2 constant region polypeptide that is fused to the hinge region
CC	polypeptide, and an immunoglobulin heavy chain CH3 constant region
CC	polypeptide that is fused to the CH2 constant region polypeptide. The

hinge region polypeptide comprises: a wild-type human IgG1 immunoglobulin hinge region polypeptide; a mutated human IgG1 immunoglobulin hinge region polypeptide, derived from (a) having 3 or more cysteine residues; where the mutated human IgG1 immunoglobulin hinge region polypeptide contains 2 cysteine residues, where the first cysteine is not mutated; a mutated human IgG1 immunoglobulin hinge region polypeptide, derived from (a) having 3 or more cysteine residues, where the mutated human IgG1 immunoglobulin hinge region polypeptide contains no more than one cysteine residue; and a mutated human IgG1 immunoglobulin hinge region polypeptide, derived from (a) having 3 or more cysteine residues; where the mutated human IgG1 immunoglobulin hinge region polypeptide contains no cysteine residues. The binding domain-immunoglobulin fusion protein is capable of at least one immunological activity comprising antibody dependent cell-mediated cytotoxicity (ADCC) and complement fixation. The binding domain polypeptide is capable of specifically binding to an antigen. Also included are an isolated polynucleotide encoding the binding domain-immunoglobulin fusion protein, a recombinant expression construct comprising the polynucleotide (operably linked to a promoter), a host cell transformed or transfected with a recombinant expression construct, producing the binding domain-immunoglobulin fusion protein, a pharmaceutical composition comprising the binding domain-immunoglobulin fusion protein or polynucleotide and a carrier, and treating a subject having or suspected of having a malignant condition or a B-cell disorder. The binding domain-immunoglobulin fusion protein is useful for treating a subject having or suspected of having a malignant condition or a B-cell disorder, e.g. melanoma, carcinoma or sarcoma, Rheumatoid arthritis, myasthenia gravis, Grave's disease, Type I diabetes mellitus, multiple sclerosis or autoimmune disease. The present sequence is a binding domain-immunoglobulin fusion protein-associated protein sequence. Note: The sequence data for this patent formed part of the printed specification and is also available in electronic format directly from USPRO at [seqdata.uspro.gov/sequence.html?DocID=20030118592](http://seqdata.uspro.gov/sequence.html?DocID=20030118592). The authors have not identified the sequences in the printed specification by their SEQ ID number therefore none of the sequences can be explicitly identified.

**SQ** Sequence 458 AA;

Query Match	87.4%;	Score 2032;	DB 7;	Length 458;
-------------	--------	-------------	-------	-------------

```

Local similarity      85.28;      E-value: NO: 1.0E-120;
Matches  405;  Conservative  2;  Mismatches  17;  Indels  30;  Gaps  2;

```

QY	1	MNRGVPFPHLLVYVQLALLPATQGNKVLLGKGGDTVELTCTASQOKSIQFPMKNSNOIK	60
Db	1	MNRGVFPFHLLVLTQALLPATQGNKVLLGKGGDTVELTCTASQOKSIQFPMKNSNOIK	60
QY	61	ILGNQGSFLTJGPKSKLNDRADSRSLMDQGNFPLIIKMLKIEDSPYICEVEDQKEEYOL	120
Db	61	ILGNQGSFLTJGPKSKLNDRADSRSLMDQGNFPLIIKMLKIEDSPYICEVEDQKEEYOL	120
QY	121	LVFGITANSDFHLLQOGOSLTVLTSPSPSSSVQCRSPRGKNIQGGKTLTSVQLBELQDSG	180
Db	121	LVFGITANSDFHLLQOGOSLTVLTSPSSSVQCRSPRGKNIQGGKTLTSVQLBELQDSG	180
QY	181	TWCTCTVLOQNKKEVEPKIDIVLAFQKASSIYKKEGEOVESFPPLAFVEXKLTGSGELMW	240
Db	181	TWCTCTVLOQNKKEVEPKIDIVLAFQKASSIYKKEGEOVESFPPLAFVEXKLTGSGELMW	240
QY	241	QAEKASSSKSIITFDLKNKEVSVKRVTPDPLQOMGKPLPHLTLTQALPQVAGSGNLTIA	300
Db	241	QAEKASSSKSIITFDLKNKEVSVKRVTPDPLQOMGKPLPHLTLTQALPQVAGSGNLTIA	300
QY	301	LEATYTKLHOEVNLVVMRATOLQKNLTCCEVMGPTSPKMLSLKENKEKVSREKRPVW	360
Db	301	LEATYTKLHOEVNLVVMRATOLQKNLTCCEVMGPTSPKMLSLKENKEKVSREKRVWV	360
QY	361	LNPEAGMWOCLLSSGOVLLESNKVLTPTWSTPVPYPRASALPAPPTGSLPDPQTASALP	420
Db	361	LNPEAGMWOCLLSSGOVLLESNKVLTPTWSTPVP-----	396
QY	421	DPAPASALPALAVISFLTGLTGLGV-ACVTLARTR	453
Db	397	-----MALITVGVAGLLTFLGLTGFFCVRCRHR	425

RESULT 40  
 ADE57489  
 ID ADE57489 standard; protein; 458 AA.  
 AC  
 XX ADE57489;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Human Protein P01730, SEQ ID NO 3351.  
 XX  
 KW Human; pain; neuronal tissue; gene therapy;  
 KW spinal segmental nerve injury; chronic constriction injury; CCI;  
 KW spared nerve injury; SNI; Chung.  
 XX  
 OS Homo sapiens.  
 PN WO2003016475-A2.  
 XX  
 PD 27-FEB-2003.  
 XX  
 PF 14-AUG-2002; 2002WO-US025765.  
 XX  
 PR 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX  
 PA (GEHO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.  
 XX  
 PI Woolf C, D'Urso D, Befort K, Coeigan M;  
 DR WPI: 268312/26.  
 DR GENBANK; P01730.  
 XX  
 PT New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 XX  
 PS Claim 1; Page; 1017pp; English.  
 XX  
 CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPo at  
 CC ftp.wipo.int/pub/published\_pcl\_sequences.  
 XX  
 XX Sequence 458 AA;  
 XX

QY	1	MMRGVPEPHLLVLTQALLPAATQGNKVLGKKGDVVELTCTASQKKSIOFHMKNSNOJK	60
Db	1	MNRGVPEPHLLVLTQALLPAATQGNKVLGKKGDVVELTCTASQKKSIOFHMKNSNOJK	60
QY	61	ILGNQGSFLTKPGSPKLNDRADSRRLMDQGNFPLIIKNLIKIEDSDVYICEVEDQKEVOL	120
Db	61	ILGNQGSFLTKPGSKLNDRADSRRLMDQGNFPLIIKNLIKIEDSDVYICEVEDQKEVOL	120
QY	121	LVFGITANSDPHTLLOGOSITLTLESPPGSSPSVOCRSPRGKNIQGGKITLSVSQLEIODSG	180
Db	121	LVFGITANSDPHTLLOGOSITLTLESPPGSSPSVOCRSPRGKNIQGGKITLSVSQLEIODSG	180
QY	181	TWTCIVLONOKKVEFKIDIVLVAPOKASSIVYKKEGEOVEFSFPLAFVTEKLTGSGELMW	240
Db	181	TWTCIVLONOKKVEFKIDIVLVAPOKASSIVYKKEGEOVEFSFPLAFVTEKLTGSGELMW	240
QY	241	QAEARSSSKSWITFDLKNKKEVSVKVETODPKLOMGKPLPLHTLPQALPOYAGSGNLTIA	300
Db	241	QAEARSSSKSWITFDLKNKKEVSVKVETODPKLOMGKPLPLHTLPQALPOYAGSGNLTIA	300
QY	301	LEAKTGKILHOENLVVMBRATOLQKULTECVMGPTSPKMLSLKLENKAKVSRREKPVWY	360
Db	301	LEAKTGKILHOENLVVMBRATOLQKULTECVMGPTSPKMLSLKLENKAKVSRREKPVWY	360
QY	361	LNPEAGMVOCLLSDSGOVLLESNIVLPTWSTPVPBPASALPAPPTGSAALPDQOTASALP	420
Db	361	LNPEAGMVOCLLSDSGOVLLESNIVLPTWSTPVPBPASALPAPPTGSAALPDQOTASALP	420
QY	421	DPBASALPAPALAVISFLIGLGLGV-ACYLATR	453
Db	397	-----MALIVLGGVAGLLIFGLGIGFCVCRCHR	425

RESULT 41	
ADAA44807	
ID	ADAA44807 standard; protein; 473 AA.
XX	
AC	
XX	
ADAA44807;	
DT	
04-DEC-2003	(first entry)
XX	
DE	CD4/TcR CD3epsilon chain chimeric protein CD4epsilon15, SEQ ID NO:2.
XX	
KW	HIV-1 infection; human immunodeficiency virus-1; CD4+ cell; chimeric CD4
KW	endoplasmic reticulum; ER retention; envelope protein gp160;
KW	T cell receptor CD3epsilon chain; C-terminal domain; CD4epsilon15;
KM	gene therapy; human; receptor.
XX	
OS	Chimeric.
OS	Homo sapiens.
XX	
PH	
FT	Key
FT	Protein
FT	Location/Qualifiers
FT	1..458
FT	/label= CD4
FT	Region
FT	459..473
FT	/note="Part of the C-terminal domain of the T cell
FT	receptor CD3epsilon chain"
XX	
FM	
WM2003076468-A1.	
XX	
PD	
18-SEP-2003.	
XX	
PE	14-MAR-2003; 2003MO-ES000120.
XX	
PR	14-MAR-2002; 2002ES-00000615.
XX	
PA	(CNSJ ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
XX	
PI	Alarcon Sanchez BJ, San Jose Martinez ME, Zaldivar Notario I,
XX	Gomez Buendia M,
DR	WPI; 2003-779059/73.

DR N-PSDB; ADA44806.  
XX Composition for treating or preventing human immune deficiency virus,  
PT comprises CD4 chimeric protein having a protective effect in trans, or  
PT related nucleic acid.  
XX Claim 5; Page 33-35; 43pp; Spanish.  
XX  
CC The invention relates to a composition for the treatment or prevention of  
CC human immunodeficiency virus-1 (HIV-1) infection. The composition  
CC comprises CD4+ cells that have been transduced with a vector that encodes  
CC a chimeric CD4 molecule which is capable of being retained in the  
CC endoplasmic reticulum (ER). The invention also encompasses the use of a  
CC soluble protein factor produced by CD4+ cells that have been transduced  
CC with a vector encoding a chimeric CD4 protein; and the use of an  
CC expression system encoding a chimeric CD4 protein. The ER-localised  
CC chimeric CD4 molecule binds to the HIV-1 envelope protein gp160,  
CC resulting in HIV-1 retention in the ER and thereby preventing viral  
CC replication. In a specific embodiment, the chimeric CD4 molecule  
CC comprises CD4 fused to 15 amino acids of the C-terminal domain of the T  
CC cell receptor CD3epsilon chain; this chimeric CD4 molecule is designated  
CC CD4epsilon15 (ADA44807). A known chimeric CD4 of similar structure but  
CC containing only 10 amino acids from CD3epsilon can also be used.  
CC Compositions of the invention have an in trans effect on the replication  
CC of HIV-1, and may be used to treat and prevent HIV-1 infection. The  
CC present sequence represents the chimeric CD4 molecule CD4epsilon15, which  
CC is specifically claimed for use in compositions of the invention.  
XX  
SQ Sequence 473 AA;

Query Match 87.4%; Score 2032; DB 7; Length 473;  
Best Local Similarity 89.2%; Pred. No. 1.6e-130;  
Matches 405; Conservative 2; Mismatches 17; Indels 30; Gaps 2;

QY 1 MNRGVPFRHLIVLQALPAATQGNKVVLGKGDVTELTCTASOKKSIOFHMKNNOIK 60  
DB 1 MNRGVPFRHLIVLQALPAATQGNKVVLGKGDVTELTCTASOKKSIOFHMKNNOIK 60  
QY 61 ILNGQGSFLTKGSPKLNDRADSRSLMDQGNFPLIINKLIEDSTYICEVEDQKEEVOL 120  
DB 61 ILNGQGSFLTKGSPKLNDRADSRSLMDQGNFPLIINKLIEDSTYICEVEDQKEEVOL 120  
QY 121 LVFGLTANSPTHLLQOGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSG 180  
DB 121 LVFGLTANSPTHLLQOGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSG 180  
QY 181 TWTCTVLOQNKKEFKIDIVLAFQKASSIVYKKEGQVEFSPFLATFVEKLTGSGELMW 240  
DB 181 TWTCTVLOQNKKEFKIDIVLAFQKASSIVYKKEGQVEFSPFLATFVEKLTGSGELMW 240  
QY 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKPLPHLTLPQALPOYAGSGNLTILA 300  
DB 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKPLPHLTLPQALPOYAGSGNLTILA 300  
QY 301 LEAKTGKHOEVNLVVMRATQLOKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPYWV 360  
DB 301 LEAKTGKHOEVNLVVMRATQLOKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPYWV 360  
QY 361 LNPEAGMOCCLISDSGOVLLESINIVLPTWSTPVHPRASALPAPPTGSLPDPOYASALP 420  
DB 361 LNPEAGMOCCLISDSGOVLLESINIVLPTWSTPVHPRASALPAPPTGSLPDPOYASALP 420  
QY 421 DEPASALPALAVISFLIGLIGLV-ACVLAATR 453  
DB 421 DEPASALPALAVISFLIGLIGLV-ACVLAATR 453  
QY 397 -----MALIVLGVAGVAGLFLITGLGIFCVCRKHR 425  
DB 397 -----MALIVLGVAGVAGLFLITGLGIFCVCRKHR 425

RESULT 42  
AAR20152  
ID AAR20152 standard; protein; 519 AA.  
XX AAR20152;  
XX

DT 25-MAR-2003 (revised)  
DT 31-MAR-1992 (first entry)  
XX Human CD4 sequence encoded by PATY.6.  
XX  
XX Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;  
XX acquired immune deficiency syndrome; AIDS related complex;  
XX T helper lymphocytes.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
XX Peptide 1..25  
XX /label= signal\_sequence  
XX  
XX W09118618-A.  
XX  
XX 12-DEC-1991.  
XX  
XX 25-MAY-1990; 90US-00529186.  
XX  
XX 25-MAY-1990; 90US-00529186.  
XX  
XX (BIOJ ) BIOGEN INC.  
XX  
XX Fisher RA, Hession C, Burkly LC;  
XX  
XX WPI; 1992-007200/01.  
XX  
XX N-PSDB; AAQ20327.  
XX  
XX New immuno-therapeutic human CD4 variants and derivs. - elicit AB  
XX production to HIV gp.120, useful in treating, preventing and diagnosing  
XX AIDS, ARC and HIV infections.  
XX  
XX Disclosure; Fig 28; 179pp; English.  
XX  
XX The sequence was deduced from the DNA sequence of subclone PATY.6, contg.  
XX DNA coding for the full-length human CD4. The clone was constructed from  
XX plasmids PBG178A and PBG378 (both in US8802940). The DNA can be used to  
XX express recombinant CD4 and analogues for use in diagnosis and treatment  
XX of diseases caused by infective agents whose primary targets are T4  
XX lymphocytes. See also AAR20148-R20155 and AAR21078. (Updated on 25-MAR-  
XX 2003 to correct PA field.)  
XX  
SQ Sequence 519 AA;

Query Match 87.4%; Score 2032; DB 2; Length 519;  
Best Local Similarity 89.2%; Pred. No. 1.8e-130;  
Matches 405; Conservative 2; Mismatches 17; Indels 30; Gaps 2;

QY 1 MNRGVPFRHLIVLQALPAATQGNKVVLGKGDVTELTCTASOKKSIOFHMKNNOIK 60  
DB 62 MNRGVPFRHLIVLQALPAATQGNKVVLGKGDVTELTCTASOKKSIOFHMKNNOIK 121  
QY 61 ILNGQGSFLTKGSPKLNDRADSRSLMDQGNFPLIINKLIEDSTYICEVEDQKEEVOL 120  
DB 122 ILNGQGSFLTKGSPKLNDRADSRSLMDQGNFPLIINKLIEDSTYICEVEDQKEEVOL 181  
QY 121 LVFGLTANSPTHLLQOGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSG 180  
DB 121 LVFGLTANSPTHLLQOGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSG 241  
QY 181 TWTCTVLOQNKKEFKIDIVLAFQKASSIVYKKEGQVEFSPFLATFVEKLTGSGELMW 240  
DB 181 TWTCTVLOQNKKEFKIDIVLAFQKASSIVYKKEGQVEFSPFLATFVEKLTGSGELMW 301  
QY 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKPLPHLTLPQALPOYAGSGNLTILA 300  
DB 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKPLPHLTLPQALPOYAGSGNLTILA 361  
QY 301 LEAKTGKHOEVNLVVMRATQLOKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPYWV 360  
DB 301 LEAKTGKHOEVNLVVMRATQLOKNLTCEVWGPTSPKLMLSLKLENKEAKVSKREKPYWV 421

```

QY 361 LNPEAGMOCCLSDSGVLTLESNIKVLPTWSTPVHPRASALPAPPTGSLPDPQTASALP 420
DB 422 LNPEAGMOCCLSDSGVLTLESNIKVLPTWSTPVDP----- 457
QY 421 DPPAASALPALAVISFLGLGLGV-ACVLTART 453
DB 458 -----MALIVLGAGVGLLFLIGLGFRCRHR 486

RESULT 43
AA06374
ID AA06374 standard; protein; 400 AA.
XX
AC AA06374;
XX
DT 31-OCT-2002 (revised)
DT 20-DEC-1990 (first entry)
XX
DE Truncated form of soluble T4 encoded by pBG381.
XX
KM plasmid pBG381; soluble T4 protein; AIDS; ARC; HIV.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..23
FT /label= secretory signal
FT /note= "hydrophobic"
FT Region 24..117
FT /label= extracellular
FT /note= "homology to V-regions"
FT Region 118..132
FT /label= extracellular
FT /note= "homology to J-regions"
FT Region 133..397
FT /label= extracellular
FT /note= "glycosylated"
XX
FM WO9008198-A.
XX
PD 26-JUL-1990.
XX
PF 18-JAN-1989; 89US-00300096.
XX
PR 18-JAN-1989; 89US-00300096.
XX
PA (HARD ) HARVARD COLLEGE.
XX
PI Letvin NA;
XX
DR WPI; 1990-254040/33.
DR N-PSDB; AAQ05608.
XX
PT Treating or preventing AIDS, ARC or HIV infection - by administering an
PT immunologically effective amt. of soluble T4 protein.
XX
PS Disclosure; Fig 2; 121pp; English.
XX
CC T4-encoding plasmid pBG381 was used to transform Chinese Hamster Ovary
CC cells for production of truncated T4. Soluble T4 is produced by
CC virtue of the removal of the transmembrane and cytoplasmic domains. The
CC soluble forms may be modified to increase their immunogenicity by
CC addition of an adjuvant such as incomplete Freund's adjuvant. The T4
CC interferes with HIV/T4 interaction and elicits anti-soluble T4 antibody
CC production. See also AAQ05607. (Updated on 31-OCT-2002 to add missing OS
CC field.)
XX
SQ Sequence 400 AA;
XX
Query Match 87.4%; Score 2031; DB 2; Length 400;
Best Local Similarity 99.2%; Pred. No. 1.5e-130;
Matches 395; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 NMRGVPFRHLLVLTALLPATQGNKVVLGKGGDTVELTCTASQKSIQPHWKNNOIK 60
DB 1 NMRGVPFRHLLVLTALLPATQGNKVVLGKGGDTVELTCTASQKSIQPHWKNNOIK 60
QY 61 ILGNQSFYLTGKPSKLANDRADSRRLMDQGNFPLIKLIKIEDSDTYICEVEDQKEEYQL 120
DB 61 ILGNQSFYLTGKPSKLANDRADSRRLMDQGNFPLIKLIKIEDSDTYICEVEDQKEEYQL 120
QY 121 LVFGLTANSDFHLLQGSLLTLESPPGSSPSVQCRSPRGKNIQCGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDFHLLQGSLLTLESPPGSSPSVQCRSPRGKNIQCGKTLVSQLELDQSG 180
QY 181 TWTCVTLOKQKVEFKIDIVLAFOKASSIYKKEGEQVESFPPLAFVEXKLTSGSELW 240
DB 181 TWTCVTLOKQKVEFKIDIVLAFOKASSIYKKEGEQVESFPPLAFVEXKLTSGSELW 240
QY 241 QAERASSKSWITPDLKNKEVSVKRVTDPKLQMGKPLHLTLPLQALPOYAGSGNLTILA 300
DB 241 QAERASSKSWITPDLKNKEVSVKRVTDPKLQMGKPLHLTLPLQALPOYAGSGNLTILA 300
QY 301 LEAKTGKLIHQEVNIVWMEATOLQKNLTCEVWGPTSPKMLSLKENKAKVSKREKPVW 360
DB 301 LEAKTGKLIHQEVNIVWMEATOLQKNLTCEVWGPTSPKMLSLKENKAKVSKREKAVW 360
QY 361 LNPEAGMOCCLSDSGVLTLESNIKVLPTWSTPVHPR 398
DB 361 LNPEAGMOCCLSDSGVLTLESNIKVLPTWSTPVQPM 398

RESULT 44
AA51081
ID AA51081 standard; protein; 481 AA.
XX
AC AA51081;
XX
DT 23-MAR-2000 (first entry)
XX
DE Human fusion protein CD4Pmu.
XX
KM Fusion protein; human; CD4; IgM; immunoglobulin; gp120;
KM anti-human immunodeficiency virus; CD4Pmu.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN US6004781-A.
XX
PD 21-DEC-1999.
XX
PF 04-FEB-1994; 94US-00191708.
XX
PR 22-JAN-1988; 88US-00147351.
PR 23-JAN-1989; 89US-00299596.
PR 09-JUN-1992; 93US-00896781.
PR 12-APR-1993; 93US-00057952.
XX
PA (GENO ) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
DR WPI; 2000-085792/07.
DR N-PSDB; AA244064.
XX
PT Fusion protein useful for the treatment of human immunodeficiency virus.
XX
PS Example 1; Col 49-58; 39pp; English.
XX
CC This invention describes a novel nucleic acid (I) encoding a fusion
CC protein comprising a DNA sequence encoding amino acids 1-173 of CD4 (II)
CC and a DNA sequence encoding a human immunoglobulin (Ig) heavy or light
CC chain (III). The products of the invention have anti-human
CC immunodeficiency virus (HIV) activity and are capable of binding to

```

CC gp120. The fusion protein is useful for treating human immunodeficiency  
 CC virus (HIV) or simian immunodeficiency virus (SIV). This sequence  
 CC represents the fusion protein CD4mu which is constructed from CD4 linked  
 CC to human IGM upstream of the CH2 region

XX Sequence 481 AA;

Query Match 87.4%; Score 2031; DB 3; Length 481;  
 Best Local Similarity 93.0%; Pred. No. 1,9e-130;  
 Matches 399; Conservative 4; Mismatches 10; Indels 16; Gaps 2;

QY 1 MNRGVPFRHLLVLTQALLPAATQGNKVLGKGGDTVELCTASQKSIQFHMKNNOIK 60  
 DB 1 MNRGVPFRHLLVLTQALLPAATQGNKVLGKGGDTVELCTASQKSIQFHMKNNOIK 60  
 QY 61 ILNGSGFLTKGSPKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEVQL 120  
 DB 61 ILNGSGFLTKGSPKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEVQL 120  
 QY 121 LVFGLTANSPTHLQGGSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSPTHLQGGSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 QY 181 TWCTVLQGNKVEFKIDIVLAFQKASSIYKKEGQVEFSFPLAFTVEKLTGSGELMW 240  
 DB 181 TWCTVLQGNKVEFKIDIVLAFQKASSIYKKEGQVEFSFPLAFTVEKLTGSGELMW 240  
 QY 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLPLQALPOYAGSGNLTLA 300  
 DB 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLPLQALPOYAGSGNLTLA 300  
 QY 301 LEAKTGKLGHOEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKAKVSKKEKPYWV 360  
 DB 301 LEAKTGKLGHOEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKAKVSKKEKPYWV 360  
 QY 361 LNPEAGMWQCLLSDSGOVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPOQTASALP 420  
 DB 361 LNPEAGMWQCLLSDSGOVLLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPOQTASALP 420  
 QY 421 DNPASALP 429  
 DB 421 DNPASALP 429  
 QY 406 -PKVSVFVP 413  
 DB 406 -PKVSVFVP 413

# RESULT 45

AA04032  
 ID AAR04032 standard; protein; 2037 AA.

XX AAR04032;  
 AC AAR04032;  
 XX 25-MAR-2003 (revised)  
 DT 31-OCT-2002 (revised)  
 DT 29-MAY-1990 (first entry)  
 XX Full length T4 encoded by plasmid pBG381.  
 DE Soluble T4; pBG381; anti-retroviral agent; AIDS; HIV; AZT.  
 KM Synthetic.  
 OS Synthetic.  
 XX MO8911860-A.  
 XX 14-DEC-1989.  
 PD 14-DEC-1989.  
 XX 08-JUN-1989; 89WO-US002453.  
 PF 10-JUN-1988; 88US-00204645.  
 PR 20-APR-1989; 89US-00341080.  
 XX (BIOJ ) BIOGEN NY INC.  
 PA (GEHO ) GEN HOSPITAL CORP.  
 PA (BIOJ ) BIOGEN INC.  
 PA (BIOJ ) BIOGEN INC.

XX Fieber RA, Schooley RT, Hirsch MS, Johnson VA, Walker BD;  
 PI WPI; 1990-007302/01.  
 XX N-PSDB; AA03006.  
 DR N-PSDB; AA03006.

XX Combinations of soluble T4 protein and anti-retroviral agent - having  
 PT synergistic activity in treatment and prevention of AIDS, arc and HIV  
 PT infection.

PS Disclosure; Fig 2; 100pp; English.

XX X = stop codon. The sequence was deduced from the cDNA insert of pBG183.  
 CC Soluble T4 constructs may be produced by truncating this sequence to give  
 CC fragments from position 400 to 799, removing the transmembrane and  
 CC intracytoplasmic domains whilst retaining the extracellular region  
 CC responsible for HIV binding. The sol. T4 is combined with an anti-viral  
 CC agent such as AZT. See also AA03005. (updated on 31-OCT-2002 to add  
 CC missing OS field.) (updated on 25-MAR-2003 to correct PA field.)

XX Sequence 2037 AA;

Query Match 87.4%; Score 2031; DB 2; Length 2037;  
 Best Local Similarity 99.2%; Pred. No. 1.1e-129;  
 Matches 395; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLLVLTQALLPAATQGNKVLGKGGDTVELCTASQKSIQFHMKNNOIK 60  
 DB 403 MNRGVPFRHLLVLTQALLPAATQGNKVLGKGGDTVELCTASQKSIQFHMKNNOIK 462  
 QY 61 ILNGSGFLTKGSPKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEVQL 120  
 DB 463 ILNGSGFLTKGSPKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEVQL 522  
 QY 121 LVFGLTANSPTHLQGGSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 523 LVFGLTANSPTHLQGGSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 582  
 QY 181 TWCTVLQGNKVEFKIDIVLAFQKASSIYKKEGQVEFSFPLAFTVEKLTGSGELMW 240  
 DB 583 TWCTVLQGNKVEFKIDIVLAFQKASSIYKKEGQVEFSFPLAFTVEKLTGSGELMW 642  
 QY 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLPLQALPOYAGSGNLTLA 300  
 DB 643 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLPLQALPOYAGSGNLTLA 702  
 QY 301 LEAKTGKLGHOEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKAKVSKKEKPYWV 360  
 DB 703 LEAKTGKLGHOEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKAKVSKKEKPYWV 762  
 QY 361 LNPEAGMWQCLLSDSGOVLLESNIKVLPTWSTPVHPRA 398  
 DB 763 LNPEAGMWQCLLSDSGOVLLESNIKVLPTWSTPVHPMA 800

# RESULT 46

AA07641  
 ID AAR07641 standard; protein; 2050 AA.

XX AAR07641;  
 AC AAR07641;  
 XX 31-OCT-2002 (revised)  
 DT 20-DEC-1990 (first entry)  
 XX Deduced sequence of pBG381 comprising truncated T4 glycoprotein.  
 DE plasmid pBG381; soluble T4 protein; AIDS; ARC; HIV.  
 KM Synthetic.  
 OS Synthetic.  
 XX Key Location/Qualifiers  
 FH 403..803  
 FT /label= truncated soluble T4 glycoprotein

XX MO9008198-A.  
 PN 26-JUL-1990.  
 PD 18-JAN-1989; 89US-00300096.  
 PF 18-JAN-1989; 89US-00300096.  
 PR (HARD ) HARVARD COLLEGE.  
 PA Letvin NA;  
 PI MPI: 1990-254040/33.  
 PS N-PSDB; AAQ05608.  
 DR Treating or preventing AIDS, ARC or HIV infection - by administering an immunologically effective amt. of soluble T4 protein.  
 PT Disclousure; Fig 2; 121pp; English.  
 XX  
 XX Entire sequence translation of plasmid pBG381 used to transform Chinese  
 CC Hamster Ovary cells for the production of soluble truncated T4.  
 CC Transmembrane and cytoplasmic domain-encoding regions are deleted from  
 CC the T4 CDS to encode a truncated protein. The soluble forms may be  
 CC modified to increase their immunogenicity by addition of an adjuvant such  
 CC as incomplete Freund's adjuvant. The T4 interferes with HIV/T4  
 CC interaction and elicits anti-soluble T4 antibody prodn. See also  
 CC AAQ05607. (updated on 31-OCT-2002 to add missing OS field.)  
 CC  
 SQ Sequence 2050 AA;

Query Match 87.4%; Score 2031; DB 2; Length 2050;  
 Best Local Similarity 99.2%; Pred. No. 1.1e-129;  
 Matches 395; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 NMRGVPFRHLVLVQLALLPAATQGNKVVLGKGGDTVELTCTASQKSIQPHMKNSNOIK 60  
 DB 403 NMRGVPFRHLVLVQLALLPAATQGNKVVLGKGGDTVELTCTASQKSIQPHMKNSNOIK 462  
 QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKLIKIEDSDTYICEVEDQKEEYQL 120  
 DB 463 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKLIKIEDSDTYICEVEDQKEEYQL 522  
 QY 121 LVFGLTANSPTHLQGGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 523 LVFGLTANSPTHLQGGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 582  
 QY 181 TWTCTVLQOKKVEFFKIDIVLAFQKASSIVYKKEGEQVEFSPFLAFTYEXKITGSGELMW 240  
 DB 583 TWTCTVLQOKKVEFFKIDIVLAFQKASSIVYKKEGEQVEFSPFLAFTYEXKITGSGELMW 642  
 QY 241 QAERASSSSSWITFPDKNKEVSVKRVTOPPKLQMGKCLPLHLITPOLPOVAGSGNLTIA 300  
 DB 643 QAERASSSSSWITFPDKNKEVSVKRVTOPPKLQMGKCLPLHLITPOLPOVAGSGNLTIA 702  
 QY 301 LEAKTGKGLHQBVLVVMRAATQLQKNLTCEVWGPTSPKMLSLKLENKAKVSKREKPVVY 360  
 DB 703 LEAKTGKGLHQBVLVVMRAATQLQKNLTCEVWGPTSPKMLSLKLENKAKVSKREKAVVY 762  
 QY 361 LNPEAGMOCCLSDSGVLLSENIKYLPTWSTPVPYHRA 398  
 DB 763 LNPEAGMOCCLSDSGVLLSENIKYLPTWSTPVPQWMA 800

RESULT 47  
 AAR13491  
 ID AAR13491 standard; proteoin; 458 AA.  
 XX AAR13491;  
 AC AAR13491;  
 XX 25-MAR-2003 (revised)  
 DT 30-OCT-1991 (first entry)

XX Human CD4 encoded by pJOD.sCD4.Y187.Snab1 and p170.2.  
 DE C4bp; gp120; HIV; T lymphocyte; fusion protein.  
 XX Homo sapiens.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..25  
 FT /label= signal\_peptide  
 FT Domain 26..132  
 FT /label= Ig-related  
 FT /note= "extracellular"  
 FT Disulfide-bond 41..109  
 FT Domain 133..202  
 FT /label= Ig-related  
 FT /note= "extracellular"  
 FT Disulfide-bond 155..184  
 FT Domain 203..318  
 FT /label= Ig-related  
 FT /note= "extracellular"  
 FT Domain 319..395  
 FT /label= Ig-related  
 FT /note= "extracellular"  
 FT Disulfide-bond 328..370  
 FT Region 396..416  
 FT /label= transmembrane  
 FT Domain 417..456  
 FT /label= cytoplasmic

MO9111461-A.  
 08-AUG-1991.  
 26-JAN-1990; 90US-00470888.  
 26-JAN-1990; 90US-00470888.  
 PASEK MP, Winkler G, Liu TR;  
 (BIOJ ) BIOGEN INC.  
 MPI: 1991-252613/34.  
 N-PSDB; AAQ13243.  
 New C4 binding protein fusion proteins and DNA encoding them - comprise  
 assemblies of C4bp monomers linked to functional moiety, e.g. AZT, useful  
 as delivery vehicles in diagnosis and therapy.  
 Example 3; Fig 3; 105pp; English.  
 This is the preferred CD4 sequence for use in the construction of fusion  
 proteins with C4-binding protein. Truncated, soluble versions of CD4 can  
 also be used. The C4bp-CD4 fusion protein may be useful to target AZT or  
 similar anti-retroviral agent to HIV-infected cells. See AAQ13242-51.  
 (Updated on 25-MAR-2003 to correct PA field.)  
 SQ Sequence 458 AA;

Query Match 87.3%; Score 2030; DB 2; Length 458;  
 Best Local Similarity 89.2%; Pred. No. 2.1e-130;  
 Matches 405; Conservative 2; Mismatches 17; Indels 30; Gaps 2;  
 QY 1 NMRGVPFRHLVLVQLALLPAATQGNKVVLGKGGDTVELTCTASQKSIQPHMKNSNOIK 60  
 DB 1 NMRGVPFRHLVLVQLALLPAATQGNKVVLGKGGDTVELTCTASQKSIQPHMKNSNOIK 60  
 QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKLIKIEDSDTYICEVEDQKEEYQL 120  
 DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKLIKIEDSDTYICEVEDQKEEYQL 120  
 QY 121 LVFGLTANSPTHLQGGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180

```

Db      121 LVFGLTANSDTHLLQGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDPSG 180
Qy      181 |TWTCVTLNOKKVEFKIDIVVLAFOKASSIVYKKEGEVSPPLAFTVEKLTSSGELMW 240
      181 |TWTCVTLNOKKVEFKIDIVVLAFOKASSIVYKKEGEVSPPLAFTVEKLTSSGELMW 240
Qy      241 QAERASSKSMITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGNLTLLA 300
      241 QAERASSKSMITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGNLTLLA 300
Db      301 LEAKTGKLGHOEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKENKEAKVSKKEKPVWV 360
      301 LEAKTGKLGHOEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKENKEAKVSKKEKPVWV 360
Qy      361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPHPRASALPAPPTGSALPDPTASALP 420
      361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPHPRASALPAPPTGSALPDPTASALP 420
Db      421 DPPASALPALAVISFLGLGLGV-ACVLARTR 453
Qy      421 DPPASALPALAVISFLGLGLGV-ACVLARTR 453
      397 -----MALIVLGVGVALLLFTGLGIFFCVRCRHR 425

```

## RESULT 48

AAP93011 standard; protein; 481 AA.

```

ID      AAP93011 standard; protein; 481 AA.
AC      AAP93011;
XX      25-MAR-2003 (revised)
XX      03-AUG-1992 (first entry)
DE      Genetic construct which encodes CD4 linked to human IgM at the Pst site
DE      upstream of the CH2 region (fusion protein CD4Pmu).
XX      Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy;
XX      diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.
OS      Homo sapiens.
XX      EP325262-A.
XX      26-JUL-1989.
XX      20-JAN-1989; 89EP-00100913.
XX      22-JAN-1988; 88US-00147351.
XX      (GEHO ) GEN HOSPITAL CORP.
XX      Seed B;
XX      WPI, 1989-214472/30.
XX      N-PSDB; AAN90359.
DR      Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV
PT      infections or detecting HIV or SIV in sample.
XX      Example; Table 4, Page 41-47; 68pp; English.

```

The fusion protein genes of the invention pref. comprises cDNA sequences which encode CD4 or a fragment which binds gp120 ligated to an expression plasmid which encodes an antibody in which the variable region of the gene has been deleted (see WO87-02671). The CD4 portion of the fusion protein may comprise the complete CD4 sequence, the 370 AA extracellular region and the membrane spanning domain, or the extracellular region. The Ig heavy chain is pref. from IgM, IgG1 or IgG3. The following are specifically claimed: fusion proteins CD4H1ambda1, CD4muu, CD4Pmu, CD4E1ambda1 (No. 67610), PC04P1ambda (No. 67609) and PC04E1ambda1 (No. 67610). (Updated on 25-MAR-2003 to correct PA field.)

Query Match 87.3%; Score 2030; DB 1; Length 481;  
 Best Local Similarity 93.0%; Pred. No. 2,2e-130;  
 Matches 399; Conservative 4; Mismatches 10; Indels 16; Gaps 2;

```

Qy      1 MNRGVPRHLLVQLALLPAATQGNRYVIGKKGDVVELCTASQKSSIQFHMNSQIK 60
      1 MNRGVPRHLLVQLALLPAATQGNRYVIGKKGDVVELCTASQKSSIQFHMNSQIK 60
Db      61 ILNGQSFLLKGPSKLNDRADSRSLMDQGNPLIINKLIEBDDTYICEVEDQKEVQL 120
      61 ILNGQSFLLKGPSKLNDRADSRSLMDQGNPLIINKLIEBDDTYICEVEDQKEVQL 120
Qy      121 LVFGLTANSDTHLLQGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDPSG 180
      121 LVFGLTANSDTHLLQGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDPSG 180
Db      181 |TWTCVTLNOKKVEFKIDIVVLAFOKASSIVYKKEGEVSPPLAFTVEKLTSSGELMW 240
      181 |TWTCVTLNOKKVEFKIDIVVLAFOKASSIVYKKEGEVSPPLAFTVEKLTSSGELMW 240
Qy      241 QAERASSKSMITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGNLTLLA 300
      241 QAERASSKSMITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGNLTLLA 300
Db      241 QAERASSKSMITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGNLTLLA 300
      241 QAERASSKSMITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGNLTLLA 300
Qy      301 LEAKTGKLGHOEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKENKEAKVSKKEKPVWV 360
      301 LEAKTGKLGHOEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKENKEAKVSKKEKPVWV 360
Db      361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPHPRASALPAPPTGSALPDPTASALP 420
      361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPHPRASALPAPPTGSALPDPTASALP 420
Qy      421 DPPASALPALAVISFLGLGLGV-ACVLARTR 453
      421 DPPASALPALAVISFLGLGLGV-ACVLARTR 453
      406 -PKVSVFVP 413

```

## RESULT 49

AAP93012 standard; protein; 614 AA.

```

ID      AAP93012 standard; protein; 614 AA.
AC      AAP93012;
XX      25-MAR-2003 (revised)
XX      03-AUG-1992 (first entry)
DE      Genetic construct which encodes CD4 linked to human IgG1 at the BstI site
DE      downstream from the hinge region (fusion protein CD4B1ambda1).
XX      Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy;
XX      diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.
OS      Homo sapiens.
XX      EP325262-A.
XX      26-JUL-1989.
XX      20-JAN-1989; 89EP-00100913.
XX      22-JAN-1988; 88US-00147351.
XX      (GEHO ) GEN HOSPITAL CORP.
XX      Seed B;
XX      WPI, 1989-214472/30.
XX      N-PSDB; AAN90360.
DR      Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV
PT      infections or detecting HIV or SIV in sample.
XX      Example; Table 5, Page 48-55; 68pp; English.

```

The fusion protein genes of the invention pref. comprises cDNA sequences which encode CD4 or a fragment which binds gp120 ligated to an expression plasmid which encodes an antibody in which the variable region of the gene has been deleted (see WO87-02671). The CD4 portion of the fusion protein may comprise the complete CD4 sequence, the 370 AA extracellular region and the membrane spanning domain, or the extracellular region. The Ig heavy chain is pref. from IgM, IgG1 or IgG3. The following are specifically claimed: fusion proteins CD4H1ambda1, CD4muu, CD4Pmu, CD4E1ambda1 (No. 67610), PC04P1ambda (No. 67609) and PC04E1ambda1 (No. 67610). (Updated on 25-MAR-2003 to correct PA field.)



XX The fusion protein genes of the invention pref. comprises cDNA sequences  
 CC which encode CD4 or a fragment which binds gp120 ligated to an expression  
 CC plasmid which encodes an antibody in which the variable region of the  
 CC gene has been deleted (see WO87-02671). The CD4 portion of the fusion  
 CC protein may comprise the complete CD4 sequence, the 370 AA extracellular  
 CC region and the membrane spanning domain, or the extracellular region. The  
 CC Ig heavy chain is pref. from IGM, IgG1 or IgG3. The following are  
 CC specifically claimed: fusion proteins CD4H1mbd1, CD4Mmu, CD4Pmu,  
 CC CD4H1mbd1, and CD4Mmu (No. 67608), CD4P1mbd1 (No. 67609) and  
 CC PCDA1mbd1 (No. 67610). (Updated on 25-MAR-2003 to correct PA field.)  
 SQ Sequence 614 AA;

Query Match 87.1%; Score 2026; DB 1; Length 614;  
 Best Local Similarity 96.1%; Pred. No. 5.7e-130;  
 Matches 397; Conservative 1; Mismatches 13; Indels 2; Gaps 1;

QY 1 MNRGVPFRHLVLVQLALPPATQGNKVLGKKGDTVELCTASQKSIQFMKNSNOIK 60  
 DB 1 MNRGVPFRHLVLVQLALPPATQGNKVLGKKGDTVELCTASQKSIQFMKNSNOIK 60  
 QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIINKLKIEDSTYICEVEDQKEVQL 120  
 DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIINKLKIEDSTYICEVEDQKEVQL 120  
 QY 121 LVFGLTANSDTHLQGSITLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSDTHLQGSITLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 QY 181 TWTCVTLQONQKVEKIDIVLAFQKASSIYKKEGEQVEFSFPLAFTVEKLTGSGELMW 240  
 DB 181 TWTCVTLQONQKVEKIDIVLAFQKASSIYKKEGEQVEFSFPLAFTVEKLTGSGELMW 240  
 QY 241 QAERASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGLTLA 300  
 DB 241 QAERASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGLTLA 300  
 QY 301 LEAKTGKLGHOEVNLVVMRAATOLQKNLTCEVWGPTSPKMLSLKLENKAKVSKREKPYVW 360  
 DB 301 LEAKTGKLGHOEVNLVVMRAATOLQKNLTCEVWGPTSPKMLSLKLENKAKVSKREKPYVW 360  
 QY 361 LNPEAGMOCCLSDSGQVLTLESNIKVLPTWSTPVH--PRASALPAPETGSLALP 411  
 DB 361 LNPEAGMOCCLSDSGQVLTLESNIKVLPTWSTPVHADPEAPELIGSPVFLFP 413

RESULT 50

AAAG79087  
 ID AAG79087 standard; protein; 458 AA.

AC AAG79087;

DT 10-DEC-2001 (first entry)

DE Amino acid sequence of human CD4 protein.

XX Human; receptor; DC-SIGN; dendritic cell; T lymphocyte; HIV; gp120;  
 KW C-type lectin; ICAM3; HIV entry; T cell; macrophage; HIV infection; CD4.

OS Homo sapiens.

PN WO200164752-A2.

PD 07-SEP-2001.

PF 28-FEB-2001; 2001WO-US006322.

PR 02-MAR-2000; 2000US-00517605.

PA (UYNV ) UNIV NEW YORK STATE.

XX (UYN1-) UNIV NIJMEGEN.

PI Littman DR, Kwon D, Van Kooyk Y, Geijtenbeek T;  
 XX WPI; 2001-602565/68.

PT An antibody for the treatment or prevention of HIV-infection comprises a  
 PR gp120 portion which binds to DC-SIGN or is exposed upon gp120 binding of  
 CC DC-SIGN due to concomitant conformational change.

PS Disclosure; Page 115-116; 131pp; English.

CC The specification describes an antibody which is specific for an  
 CC antigenic fragment of gp120. This antigenic fragment binds to DC-SIGN or  
 CC is exposed upon gp120 binding of DC-SIGN due to concomitant  
 CC conformational change. DC-SIGN is a receptor that is specifically  
 CC expressed on dendritic cells and facilitates infection of T lymphocytes  
 CC with HIV. DC-SIGN is identical to a HIV-1 gp120-binding C-type lectin. DC  
 CC -SIGN binds ICAM-3 (which is expressed constitutively on T lymphocytes)  
 CC with high affinity. The antibody of the invention inhibits the trans  
 CC enhancement of HIV entry into a T cell or macrophage facilitated by  
 CC dendritic cells. The antibody is useful to treat or prevent HIV  
 CC infection. The present sequence represents a human CD4 protein

SQ Sequence 458 AA;

Query Match 87.1%; Score 2024; DB 4; Length 458;  
 Best Local Similarity 89.0%; Pred. No. 5.4e-130;  
 Matches 404; Conservative 2; Mismatches 18; Indels 30; Gaps 2;

QY 1 MNRGVPFRHLVLVQLALPPATQGNKVLGKKGDTVELCTASQKSIQFMKNSNOIK 60  
 DB 1 MNRGVPFRHLVLVQLALPPATQGNKVLGKKGDTVELCTASQKSIQFMKNSNOIK 60  
 QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIINKLKIEDSTYICEVEDQKEVQL 120  
 DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIINKLKIEDSTYICEVEDQKEVQL 120  
 QY 121 LVFGLTANSDTHLQGSITLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSDTHLQGSITLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 QY 181 TWTCVTLQONQKVEKIDIVLAFQKASSIYKKEGEQVEFSFPLAFTVEKLTGSGELMW 240  
 DB 181 TWTCVTLQONQKVEKIDIVLAFQKASSIYKKEGEQVEFSFPLAFTVEKLTGSGELMW 240  
 QY 241 QAERASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGLTLA 300  
 DB 241 QAERASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPOALPOYAGSGLTLA 300  
 QY 301 LEAKTGKLGHOEVNLVVMRAATOLQKNLTCEVWGPTSPKMLSLKLENKAKVSKREKPYVW 360  
 DB 301 LEAKTGKLGHOEVNLVVMRAATOLQKNLTCEVWGPTSPKMLSLKLENKAKVSKREKPYVW 360  
 QY 361 LNPEAGMOCCLSDSGQVLTLESNIKVLPTWSTPVHPRASALPAPETGSLALP 420  
 DB 361 LNPEAGMOCCLSDSGQVLTLESNIKVLPTWSTPVHPRASALPAPETGSLALP 420  
 QY 421 DPPAASALPALAVISFLIGLGV-ACVLARTR 453  
 DB 397 -----MALIVLGVAGLILFTGLGIFCVRCHRH 425

RESULT 51

AAAP93506  
 ID AAP93506 standard; protein; 394 AA.

AC AAP93506;

DT 25-MAR-2003 (revised)

DT 02-JUN-1990 (first entry)

DE Derived sequence of soluble T4 lymphocyte surface protein (sT4).

XX Soluble T4 lymphocyte surface protein; sT4; AIDS therapy; AIDS diagnosis.

```

XX OS Homo sapiens.
XX FH Key
XX FT Protein 26..394
XX FT Misc-difference 26..26
FT /note= "When sequence was determined by amino acid
FT sequencing, this residue was Lys."
FT Region
FT 27..45
FT /note= "These residues are identical to those determined
FT by amino acid sequencing"
XX PN EP313377-A.
XX PD 26-APR-1989.
XX XX
XX PF 21-OCT-1988; 88EP-00309907.
XX PR 23-OCT-1987; 87US-00112800.
XX PA (SMIK ) SMITHKLINE BECKMAN CORP.
XX PA (SMIK ) SMITHKLINE BECKMAN CORP.
XX PI Deen KC, Folenawass GM, Inacker RH, Sweet RW,
XX DR WPI; 1989-124209/17.
XX DR N-PSDB; AAN90763.
XX PT Purifying soluble recombinant T4 lymphocyte surface protein - from cell
XX PT culture by adsorption on cation exchanger, elution and treatment with
XX PT anion exchanger.
XX PS Disclosure; Fig 1; 13pp; English.
XX CC The coding sequence is derived from the published sequence of sT4. sT4 is
XX CC useful in the prevention and treatment of AIDS by inhibiting spread of
XX CC the virus. It can also be used as an inhibitor of T4+ cell function, as a
XX CC reagent for identifying inhibitors of T4+ cell interaction and to produce
XX CC diagnostic monoclonal antibodies. (Updated on 25-MAR-2003 to correct PA
XX CC field.) (Updated on 25-MAR-2003 to correct PI field.)
XX SQ Sequence 394 AA;

Query Match 86.8%; Score 2018; DB 1; Length 394;
Best Local Similarity 99.5%; Pred. No. 1,2e-129;
Matches 392; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVLTQALLPAATQGNKVVLGKGDVTELTCTASQKSIQFHMKNNSQIK 60
DB 1 MNRGVPRHLLVLTQALLPAATQGNKVVLGKGDVTELTCTASQKSIQFHMKNNSQIK 60
QY 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNPFIITIKNLKIEDSDTYICEVEDQKEEVQL 120
DB 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNPFIITIKNLKIEDSDTYICEVEDQKEEVQL 120
QY 121 LVFGLTANSPTHLLQGSILTLTLESPPGSSPSVQCRSPRGKNIQSGKTLVSQLELDQSG 180
DB 121 LVFGLTANSPTHLLQGSILTLTLESPPGSSPSVQCRSPRGKNIQSGKTLVSQLELDQSG 180
QY 121 LVFGLTANSPTHLLQGSILTLTLESPPGSSPSVQCRSPRGKNIQSGKTLVSQLELDQSG 180
DB 121 LVFGLTANSPTHLLQGSILTLTLESPPGSSPSVQCRSPRGKNIQSGKTLVSQLELDQSG 180
QY 181 TWTCTVLONOKKVERKIDIVLAFOKASSIYKKEGBOVESFPLAFTVEKLTGSGELMW 240
DB 181 TWTCTVLONOKKVERKIDIVLAFOKASSIYKKEGBOVESFPLAFTVEKLTGSGELMW 240
QY 241 QAERASSSSKSWITPFLKNKEVSVKRVTDPKLQMGKPLPHLTLPQALPOVAGSNNLTLA 300
DB 241 QAERASSSSKSWITPFLKNKEVSVKRVTDPKLQMGKPLPHLTLPQALPOVAGSNNLTLA 300
QY 301 LEAKTGKJLHOEVNVLVVRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKYSKTEKPYVW 360
DB 301 LEAKTGKJLHOEVNVLVVRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKYSKTEKPYVW 360
QY 361 LNPEAGMQCLISDSGOVLLESNIVLPTWSTPV 394
DB 361 LNPEAGMQCLISDSGOVLLESNIVLPTWSTPV 394

```

```

DB 361 LNPEAGMQCLISDSGOVLLESNIVLPTWSTPV 394

RESULT 52
AAP91922
ID AAP91922 standard; protein; 402 AA.
XX
XX AAP91922;
XX AC
XX AC AAP91922;
XX DT 25-MAR-2003 (revised)
XX DT 31-OCT-2002 (revised)
XX DT 14-MAY-1990 (first entry)
XX
XX Sequence of a secreted form of the CD4 adhesion (CD4T) polypeptide.
XX CD4 variants; CD4T; gp120; plasmid pRKC4; HIV-1; HTLV-IIIB.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH Misc-difference 25..26
XX FT /note= "signal processing site"
XX FT Misc-difference 366
XX FT /note= "other forms of CD4T terminate here"
XX FT Misc-difference 368
XX FT /note= "other forms of CD4T terminate here"
XX PN EP314317-A.
XX PD 03-MAY-1989.
XX XX
XX PF 03-OCT-1988; 88EP-00309194.
XX PR 02-OCT-1987; 87US-00104329.
XX PR 28-SEP-1988; 88US-00250785.
XX PA (GETH ) GENENTECH INC.
XX PI Capon DJ, Gregory TJ,
XX DR WPI; 1989-131855/18.
XX DR N-PSDB; AAN90777.
XX CC Compars. contg. adhesion variants - useful in therapy and diagnostics,
XX CC e.g. CD4 variants which are therapeutically useful for treating human
XX CC immuno-deficiency virus.
XX PS Disclosure; Fig 1a-1c; 36pp; English.
XX CC It may be capable of binding gp120. It may be fused with an
XX CC immunoglobulin constant domain, human transferrin, apolipoprotein,
XX CC albumin, ricin A chain or diphtheria toxin A. It may be used for
XX CC antiviral of immunomodulatory therapy particularly in treatment of HIV
XX CC infection. It may have variants by insertion, substitution of deletion in
XX CC non-functional regions. (Updated on 31-OCT-2002 to add missing OS field.)
XX CC (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to
XX CC correct PI field.)
XX SQ Sequence 402 AA;

Query Match 86.8%; Score 2017; DB 1; Length 402;
Best Local Similarity 99.7%; Pred. No. 1,4e-129;
Matches 392; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVLTQALLPAATQGNKVVLGKGDVTELTCTASQKSIQFHMKNNSQIK 60
DB 1 MNRGVPRHLLVLTQALLPAATQGNKVVLGKGDVTELTCTASQKSIQFHMKNNSQIK 60
QY 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNPFIITIKNLKIEDSDTYICEVEDQKEEVQL 120
DB 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNPFIITIKNLKIEDSDTYICEVEDQKEEVQL 120
QY 121 LVFGLTANSPTHLLQGSILTLTLESPPGSSPSVQCRSPRGKNIQSGKTLVSQLELDQSG 180
DB 121 LVFGLTANSPTHLLQGSILTLTLESPPGSSPSVQCRSPRGKNIQSGKTLVSQLELDQSG 180

```

```

Db      121 LVFGLTANSDTHLQGS/LTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Qy      181 TWTCTVLQNOKKVEFKIDIVLAFOKASSIYKKKEGEVFSFPLAFTVEKLTGSGELMW 240
Db      181 TWTCTVLQNOKKVEFKIDIVLAFOKASSIYKKKEGEVFSFPLAFTVEKLTGSGELMW 240
Qy      241 QAEKASSSSKSWITTFDLKKEVSVKRVTPDKLQMGKCLPLHLTLPOALPOYAGSGNLTLA 300
Db      241 QAEKASSSSKSWITTFDLKKEVSVKRVTPDKLQMGKCLPLHLTLPOALPOYAGSGNLTLA 300
Qy      301 LEAKTGKLGHOEVNLVVMRAATQLOKULTCVWGPTSPKMLSLKLENKEAKVSKREKPVW 360
Db      301 LEAKTGKLGHOEVNLVVMRAATQLOKULTCVWGPTSPKMLSLKLENKEAKVSKREKPVW 360
Qy      361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTP 393
Db      361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTP 393

RESULT 53
AAP94757
ID AAP94757 standard; protein; 402 AA.
AC AAP94757;
XX
XX
DT 25-MAR-2003 (revised)
DT 03-OCT-2002 (revised)
DT 28-JAN-1991 (first entry)
XX
XX
DE Sequence of a secreted form of the CD4 adhesion.
XX
XX
XX HIV; antiviral; therapy; diagnosis.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT 1..25
XX FT Peptide
XX FT Protein 26..402
XX
XX W08902922-A.
XX
XX 06-APR-1989.
XX
XX 03-OCT-1988; 88WO-US003414.
XX
XX 02-OCT-1987; 87US-00104329.
XX
XX 28-SEP-1988; 88US-00250785.
XX
XX (GETH ) GENENTECH INC.
XX
XX Capon DJ, Gregory TJ;
XX
XX WPI; 1989-114397/15.
XX
XX N-PSDB; AAN90734.
XX
XX New nucleic acid sequences encoding adhesion, esp. CD 4, variants -
XX PT partic. with trans-membrane domain inactivated or fused to other peptide,
XX PT useful esp. for treating HIV infections.
XX
XX
XX Disclosure; Fig 1a-1c; 78pp; English.
XX
XX The patent claims a nucleic acid encoding an aa sequence variant of an
XX adhesion, which is pref. a CD4 polypeptide variant modified such that its
XX transmembrane domain has been inactivated, either deleted or replaced by
XX a sequence of hydrophilic hydrophobic profile. The aa sequence variant of
XX an adhesion may also be a fusion of CD4 with a 2nd polypeptide esp. one
XX congng. a non-CD4 epitope; a signal sequence; a cpd. able to elicit a
XX humoral immune response (viral polypeptide or antigen); or a human
XX plasma protein of long plasma half-life. CD4 fusion proteins can have
XX antiviral and immunomodulatory activity and are esp. useful for treating
XX HIV infections regardless of genetic variation within the virus. They and

```

```

CC antibodies raised against them can also be used diagnostically for
CC assaying adhesion and their ligands. (Updated on 03-OCT-2002 to add
CC missing OS field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated
CC on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 402 AA;
Qy      Query Match      86.8%; Score 2017; DB 1; Length 402;
Db      Best Local Similarity 99.7%; Pred. No. 1,4e-129; Indels 0; Gaps 0;
        Matches 392; Conservative 0; Mismatches 1;
Qy      1 MNRGVPFPHLLVLVQLALLPAAQGNKVLGKGGDTVELCTASQKSIQPHMKNSNOIK 60
Db      1 MNRGVPFPHLLVLVQLALLPAAQGNKVLGKGGDTVELCTASQKSIQPHMKNSNOIK 60
Qy      61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIKLKIEDSPYICVEVDQKEBYOL 120
Db      61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIKLKIEDSPYICVEVDQKEBYOL 120
Qy      121 LVFGLTANSDTHLQGS/LTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Db      121 LVFGLTANSDTHLQGS/LTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
Qy      181 TWTCTVLQNOKKVEFKIDIVLAFOKASSIYKKKEGEVFSFPLAFTVEKLTGSGELMW 240
Db      181 TWTCTVLQNOKKVEFKIDIVLAFOKASSIYKKKEGEVFSFPLAFTVEKLTGSGELMW 240
Qy      241 QAEKASSSSKSWITTFDLKKEVSVKRVTPDKLQMGKCLPLHLTLPOALPOYAGSGNLTLA 300
Db      241 QAEKASSSSKSWITTFDLKKEVSVKRVTPDKLQMGKCLPLHLTLPOALPOYAGSGNLTLA 300
Qy      301 LEAKTGKLGHOEVNLVVMRAATQLOKULTCVWGPTSPKMLSLKLENKEAKVSKREKPVW 360
Db      301 LEAKTGKLGHOEVNLVVMRAATQLOKULTCVWGPTSPKMLSLKLENKEAKVSKREKPVW 360
Qy      361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTP 393
Db      361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTP 393

RESULT 54
AAV88328
ID AAV88328 standard; protein; 394 AA.
AC AAV88328;
XX
XX 14-JUN-2000 (first entry)
XX
XX T4 glycoprotein amino acid sequence.
XX
XX sT4; glycoprotein; human immunodeficiency virus; HIV; block binding;
XX AIDS; treatment; inhibit; cell to cell spread; infection; fusion.
XX
XX Mammalia.
XX
XX US5126433-A.
XX
XX 30-JUN-1992.
XX
XX 23-OCT-1987; 87US-00114244.
XX
XX 21-AUG-1986; 86US-00898587.
XX
XX (UYCO ) UNIV COLUMBIA NEW YORK.
XX
XX Maddon PJ, Chess L, Axel R, Weiss R, Littman DR, McDougal JS;
XX
XX WPI; 2000-348913/30.
XX
XX N-PSDB; AAA10906.
XX
XX Soluble T4 glycoprotein useful for prevention and treatment of acquired
XX immunodeficiency syndrome and for screening inhibitors of human
XX immunodeficiency viral binding.
PT

```

XX Discloure; Col 11-16; 64pp; English.  
 PS This sequence represents the full length amino acid sequence of  
 CC glycosylated sT4 glycoprotein. Human immunodeficiency virus (HIV) uses  
 CC sT4 as a target receptor on T cells. The invention relates to  
 CC glycosylated sT4 which functions by blocking the binding of HIV to T4  
 CC target cells, and can be used for the prophylaxis and treatment of AIDS  
 CC patients. Administration of sT4 effectively inhibits the cell to cell  
 CC spreading of HIV infection and also the fusion of HIV-infected T4 cells  
 CC and non-infected T4 cells. The administration of T4 alleviates several  
 CC symptoms associated with AIDS, and prevents the occurrence of new  
 CC pathological changes. The sT4 glycoprotein is useful for the prophylaxis  
 CC and treatment of patients with AIDS. It is also useful as a reagent to  
 CC identify natural, synthetic or recombinant molecules which act as  
 CC therapeutic agents or inhibitors of T4+ cell interactions and in  
 CC diagnostic assays for detection T4 proteins or molecules  
 XX Sequence 394 AA;  
 SQ

Query Match 86.7%; Score 2015; DB 3; Length 394;  
 Best Local Similarity 99.5%; Pred. No. 1.9e-129;  
 Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLIVLQALLPAAATQGNKRVLGKKGDTVELTCTASQKKSIOFHMKNNOIK 60  
 DB 1 MNRGVPFRHLIVLQALLPAAATQGNKRVLGKKGDTVELTCTASQKKSIOFHMKNNOIK 60  
 QY 61 ILNGSGFLTKGPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOL 120  
 DB 61 ILNGSGFLTKGPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOL 120  
 QY 121 LVFGLTANSPTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSPTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 QY 181 TWCTCTVLONQKKEFKIDIVLAFQKASSIYKKGBQVEFSFPLAFTVEKLTSGSELW 240  
 DB 181 TWCTCTVLONQKKEFKIDIVLAFQKASSIYKKGBQVEFSFPLAFTVEKLTSGSELW 240  
 QY 241 QABRASSSKSWITFDLKNKEVSVKRYTQDPKLGKGLPLHLTLPOLPYAGSGNLTLLA 300  
 DB 241 QABRASSSKSWITFDLKNKEVSVKRYTQDPKLGKGLPLHLTLPOLPYAGSGNLTLLA 300  
 QY 301 LEAKTGKHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKAKYSKREKPYW 360  
 DB 301 LEAKTGKHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKAKYSKREKPYW 360  
 QY 361 LNPEAGMOCCLSDSGOVLLESNIVLPTWSTPV 394  
 DB 361 LNPEAGMOCCLSDSGOVLLESNIVLPTWSTPV 394

RESULT 55  
 ADE65841 standard; protein; 458 AA.  
 XX ADE65841;  
 XX  
 XX 29-JAN-2004 (first entry)  
 DE Human CD4 receptor.  
 XX  
 XX Human; CD4 receptor; receptor; protein-protein interaction;  
 KW protein array; PDZ domain; drug target screening.  
 XX Homo sapiens.  
 OS US2003170723-A1.  
 XX PD 11-SEP-2003.  
 XX 06-MAR-2002; 2002US-00092138.  
 PF

XX 06-MAR-2002; 2002US-00092138.  
 PR (SATO/) SATO T.  
 XX  
 PA Sato T;  
 XX  
 PI WPI; 2003-852032/79.  
 DR  
 XX Preparing a protein array useful for screening drug targets comprises  
 PT depositing an array of a first protein on substrate, and applying a  
 PT second protein comprising an amino acid sequence that binds to a domain  
 PT of the first protein.  
 PS  
 XX Discloure; SEQ ID NO 25; 60pp; English.  
 CC The invention relates to a method for preparing a protein array based on  
 CC protein-protein interaction, by depositing an array of a first protein  
 CC comprising a PDZ domain on a substrate, and applying a second protein  
 CC comprising an amino acid sequence that binds to the PDZ domain of the  
 CC first protein. The method is useful for preparing protein arrays based on  
 CC biochemical protein-protein interactions. Arrays produced by this method  
 CC are useful for screening drug targets. This sequence represents the human  
 CC CD4 receptor, used in the method of the invention.  
 XX  
 XX Sequence 458 AA;  
 SQ

Query Match 86.6%; Score 2014; DB 7; Length 458;  
 Best Local Similarity 88.5%; Pred. No. 2.6e-129;  
 Matches 402; Conservative 3; Mismatches 19; Indels 30; Gaps 2;

QY 1 MNRGVPFRHLIVLQALLPAAATQGNKRVLGKKGDTVELTCTASQKKSIOFHMKNNOIK 60  
 DB 1 MNRGVPFRHLIVLQALLPAAATQGNKRVLGKKGDTVELTCTASQKKSIOFHMKNNOIK 60  
 QY 61 ILNGSGFLTKGPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOL 120  
 DB 61 ILNGSGFLTKGPSKLNDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOL 120  
 QY 121 LVFGLTANSPTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSPTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 QY 181 TWCTCTVLONQKKEFKIDIVLAFQKASSIYKKGBQVEFSFPLAFTVEKLTSGSELW 240  
 DB 181 TWCTCTVLONQKKEFKIDIVLAFQKASSIYKKGBQVEFSFPLAFTVEKLTSGSELW 240  
 QY 241 QABRASSSKSWITFDLKNKEVSVKRYTQDPKLGKGLPLHLTLPOLPYAGSGNLTLLA 300  
 DB 241 QABRASSSKSWITFDLKNKEVSVKRYTQDPKLGKGLPLHLTLPOLPYAGSGNLTLLA 300  
 QY 301 LEAKTGKHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKAKYSKREKPYW 360  
 DB 301 LEAKTGKHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKAKYSKREKPYW 360  
 QY 361 LNPEAGMOCCLSDSGOVLLESNIVLPTWSTPVQP----- 396  
 DB 421 DPEAASLPPAALAVISFLGLGCV-ACVLAARR 453  
 397 -----MALIVLGGVAGLLFLGIGIFPCVCRHR 425

RESULT 56  
 AAY39825 standard; protein; 394 AA.  
 XX AAY39825;  
 XX  
 XX 03-DEC-1999 (first entry)  
 XX Soluble human T4 protein.  
 DE

```

XX Soluble T4 protein: eT4; human: HIV; binding inhibitor: T4+ cell; AIDS;
KW vaccine; immunisation; therapy.
XX Homo sapiens.
XX US5958678-A.
XX 28-SEP-1999.
XX
XX 12-DEC-1994; 94US-00354452.
XX
XX 21-AUG-1986; 86US-00898587.
XX 11-JUN-1991; 91US-00733564.
XX 06-JUL-1992; 92US-00909021.
XX
XX (UYCO ) UNIV COLUMBIA NEW YORK.
XX
XX McDougal JS, Weiss R, Axel R, Littman DR, Maddon PJ, Chess L;
XX
XX WPI; 1999-561025/47.
XX N-PSDB; AA206594.
XX
XX Human T4 protein inhibits HIV binding to T4 cells, useful for treating
XX AIDS.
XX
XX Disclosure; Col 13-16; 58pp; English.
XX
XX This sequence represents the soluble human T4 protein of the invention.
XX The soluble human T4 protein blocks the binding of HIV to T4+ cells and
XX is therefore useful for the treatment of AIDS. Monoclonal antibodies
XX against the T4 protein may be used as vaccines for immunising subjects
XX against AIDS
XX
XX Sequence 394 AA;
SQ
Query Match      86.5%; Score 2012; DB 2; Length 394;
Best Local Similarity 99.2%; Pred. No. 3e-129;
Matches 391; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 MNRGVFPHLLVLTQALLPAATOGNKVVLGKKGDVVELTCTASOKKSIOFHMKNNOIK 60
DB 1 MNRGVFPHLLVLTQALLPAATOGKKVVLGKKGDVVELTCTASOKKSIOFHMKNNOIK 60
QY 61 ILNGGSLFKGPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILNGGSLFKGPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLQOSLTLTLESPGSSPSVQCSPPRGKNIQGGKTLVSQLELODSG 180
DB 121 LVFGLTANSDTHLQOSLTLTLESPGSSPSVQCSPPRGKNIQGGKTLVSQLELODSG 180
QY 181 TWTCTVLQOKKVEFKIDIVLAFOKASSIVYKKEGEQVEFPLAFTVEKLTGSGELMW 240
DB 181 TWTCTVLQOKKVEFKIDIVLAFOKASSIVYKKEGEQVEFPLAFTVEKLTGSGELMW 240
QY 241 QAEPRASSSSKSWITTFDLKKEVSKRVTPQPKLQMGKKLPLHLTLPOALFOVYGSGLTLTA 300
DB 241 QAEPRASSSSKSWITTFDLKKEVSKRVTPQPKLQMGKKLPLHLTLPOALFOVYGSGLTLTA 300
QY 301 LEAKTGKLEQVNLVVMRAATOLQKNTCEVMGPTSPKMLSLKLENKAKVKRRKPVNV 360
DB 301 LEAKTGKLEQVNLVVMRAATOLQKNTCEVMGPTSPKMLSLKLENKAKVKRRKPVNV 360
QY 361 LNPEAGMOCCLSDSGVLLSNIKVLPTWSTPV 394
DB 361 LNPEAGMOCCLSDSGVLLSNIKVLPTWSTPV 394
RESULT 57
AA06373
AA06373 standard; protein; 458 AA.
XX

```

```

AC AA06373;
XX
XX 31-OCT-2002 (revised)
XX 20-DEC-1990 (first entry)
XX
XX T4 encoded by plasmid p170-2.
XX
XX plasmid p170-2; soluble T4 protein; AIDS; HIV.
XX
XX Synthetic.
XX
XX Key
XX Peptide
XX 1..23
XX /label= hydrophobic/secretory signal
XX 24..117
XX /label= extracellular
XX /note= "homology to V-regions"
XX 118..132
XX /label= extracellular
XX /note= "homology to J-regions"
XX 133..397
XX /label= extracellular
XX /note= "glycosylated region"
XX 398..418
XX /label= transmembrane sequence
XX /note= "hydrophobic"
XX 419..458
XX /label= intracytoplasmic
XX /note= "very hydrophilic"
XX
XX MO9008198-A.
XX
XX 26-JUL-1990.
XX
XX 18-JAN-1989; 89US-00300096.
XX
XX 18-JAN-1989; 89US-00300096.
XX
XX (HARD ) HARVARD COLLEGE.
XX
XX Letvin NA;
XX
XX WPI; 1990-254040/33.
XX N-PSDB; AA005607.
XX
XX Treating or preventing AIDS, ARC or HIV infection - by administering an
XX immunologically effective amt. of soluble T4 protein.
XX
XX Disclosure; Fig 1; 121pp; English.
XX
XX Soluble T4 can be produced by truncating the CDS to remove the
XX transmembrane and cytoplasmic domains. The soluble forms may be modified
XX to increase their immunogenicity by addition of an adjuvant such as
XX incomplete Freund's adjuvant. The T4 interferes with HIV/T4 interaction
XX and elicits anti-soluble T4 antibody production. Soluble T4 proteins
XX include the following polypeptide fragments: amino acids 1-385, 24-385,
XX Met-24-385, 24-397, 1-400 and Met-24-400. See also AA005608. (Updated on
XX 31-OCT-2002 to add missing OS field.)
XX
XX Sequence 458 AA;
SQ
Query Match      86.5%; Score 2010; DB 2; Length 458;
Best Local Similarity 88.8%; Pred. No. 4.9e-129;
Matches 403; Conservative 2; Mismatches 19; Indels 30; Gaps 2;
QY 1 MNRGVFPHLLVLTQALLPAATOGNKVVLGKKGDVVELTCTASOKKSIOFHMKNNOIK 60
DB 1 MNRGVFPHLLVLTQALLPAATOGKKVVLGKKGDVVELTCTASOKKSIOFHMKNNOIK 60
QY 61 ILNGGSLFKGPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILNGGSLFKGPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEVOL 120

```

```

Qy 121 LVFGLTANSDTHLLOGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDPSG 180
Db 121 LVFGLTANSDTHLLOGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDPSG 180
Qy 181 TWTCVTVLONQKVEFKIDIVLAFQKASIVYKKEGEQVEFSFPLAFTVEKLTSGGELMW 240
Db 181 TWTCVTVLONQKVEFKIDIVLAFQKASIVYKKEGEQVEFSFPLAFTVEKLTSGGELMW 240
Qy 241 QAERASSKSMITFDLKNKEVSVKRVYTOPDKLQMGKKLPLHLTLPQALPOYAGSGNLTLLA 300
Db 241 QAERASSKSMITFDLKNKEVSVKRVYTOPDKLQMGKKLPLHLTLPQALPOYAGSGNLTLLA 300
Qy 301 LEAKTGKTHQEVNLVVMRATOLQKNLTCEVWGPTSPKLMSTLKENKEAKVSKREKPVW 360
Db 301 LEAKTGKTHQEVNLVVMRATOLQKNLTCEVWGPTSPKLMSTLKENKEAKVSKREKAVW 360
Qy 361 LNPEAGMWQCLLSDSGVLLSESNIKVLPWTSTPVQP----- 396
Db 361 LNPEAGMWQCLLSDSGVLLSESNIKVLPWTSTPVQP----- 396
Qy 421 DPPAASALPAALAVISFLGLGLGV-ACVLARTR 453
Db 397 -----MALIVLGGVAGLLPFLGLGIFFCVRCRHR 425

```

```

RESULT 58
AAP94703
ID AAP94703 standard; protein; 524 AA.
XX AAP94703;
AC AAP94703;
XX 25-MAR-2003 (revised)
XX DT 22-MAR-1991 (first entry)
XX DE Sequence encoded by T4 lymphocyte cDNA obtained from PBL clone lambda-203
DE -4.
XX HIV; soluble T4; immunotherapeutic; prophylactic; diagnostic; AIDS; ARC.
XX KM Homo sapiens.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Misc-difference 67 /note= "AA DESIGNATED NUMBER -23"
XX FT Misc-difference 90 /note= "AA DESIGNATED NUMBER 1"
XX FT Misc-difference 92 /note= "MATURE N-TERMINUS"
XX FT
XX PN W08901940-A.
XX PD 09-MAR-1989.
XX PF 01-SEP-1988; 88WO-US002940.
XX PR 04-SEP-1987; 87US-00094322.
XX PR 07-JAN-1988; 88US-00141649.
XX PA (BIOJ ) BIOGEN INC.
XX PI Fisher RA, Gilbert W, Sato VL, Flavell RA, Maraganore JM,
XX WPI; 1989-085519/11.
XX DR N-PSDB; AAN90642.
XX PT DNA sequences coding for soluble T4-like polypeptide(s) - used in
XX immunotherapeutic and immunosuppressive compns. and for preventing,
XX treating or detecting AIDS.
XX PS Disclosure; Fig 3; 207pp; English.
XX CC The polypeptides encoded are useful in immunotherapeutic, prophylactic
XX and diagnostic compns. They can be used to purify HIV from a sample. The
CC

```

```

CC soluble T4 protein-based compns. are useful in treating immunodeficient
CC patients suffering from diseases caused by agents whose primary targets
CC are T4+ lymphocytes. They can be used for preventing, treating or
CC detecting AIDS, ARC and HIV infection. (Updated on 25-MAR-2003 to correct
CC PR field.)
XX
SQ Sequence 524 AA;

```

```

Query Match      86.5%; Score 2010; DB 1; Length 524;
Best Local Similarity 88.8%; Pred. No. 5, 6e-129;
Matches 403; Conservative 2; Mismatches 19; Indels 30; Gaps 2;

```

```

Qy 1 MNRGVPFRHLILVQLALLPAATQGNKRVLGKGDVLTCTASQKKSIOFHMWNSNOIK 60
Db 67 MNRGVPFRHLILVQLALLPAATQGNKRVLGKGDVLTCTASQKKSIOFHMWNSNOIK 126
Qy 61 ILNGQSFLLTKGSKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEVOL 120
Db 127 ILNGQSFLLTKGSKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEVOL 186
Qy 121 LVFGLTANSDTHLLOGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDPSG 180
Db 187 LVFGLTANSDTHLLOGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDPSG 246
Qy 181 TWTCVTVLONQKVEFKIDIVLAFQKASIVYKKEGEQVEFSFPLAFTVEKLTSGGELMW 240
Db 247 TWTCVTVLONQKVEFKIDIVLAFQKASIVYKKEGEQVEFSFPLAFTVEKLTSGGELMW 306
Qy 241 QAERASSKSMITFDLKNKEVSVKRVYTOPDKLQMGKKLPLHLTLPQALPOYAGSGNLTLLA 300
Db 307 QAERASSKSMITFDLKNKEVSVKRVYTOPDKLQMGKKLPLHLTLPQALPOYAGSGNLTLLA 366
Qy 301 LEAKTGKTHQEVNLVVMRATOLQKNLTCEVWGPTSPKLMSTLKENKEAKVSKREKPVW 360
Db 367 LEAKTGKTHQEVNLVVMRATOLQKNLTCEVWGPTSPKLMSTLKENKEAKVSKREKAVW 426
Qy 361 LNPEAGMWQCLLSDSGVLLSESNIKVLPWTSTPVQP----- 462
Db 427 LNPEAGMWQCLLSDSGVLLSESNIKVLPWTSTPVQP----- 462
Qy 421 DPPAASALPAALAVISFLGLGLGV-ACVLARTR 453
Db 463 -----MALIVLGGVAGLLPFLGLGIFFCVRCRHR 491

```

```

RESULT 59
AAR07640
ID AAR07640 standard; protein; 2458 AA.
XX AAR07640;
XX AC AAR07640;
XX DT 31-OCT-2002 (revised)
XX DT 20-DEC-1990 (first entry)
XX DE Deduced protein sequence of p170-2 comprising T4.
XX DE plasmid p170-2; soluble T4 protein; AIDS; ARC; HIV.
XX KM Synthetic.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX FT Protein 400..858
XX FT /label= T4 surface glycoprotein
XX PN W09008198-A.
XX PD 26-JUL-1990.
XX PF 18-JAN-1989; 89US-00300096.
XX PR 18-JAN-1989; 89US-00300096.
XX PA (HARD ) HARVARD COLLEGE.
XX

```

```

XX      LetvIn NA;
PI      MPI: 1990-254040/33.
XX      N-PSDB; AAQ05607.
DR      Treating or preventing AIDS, ARC or HIV infection - by administering an
PT      immunologically effective amt. of soluble T4 protein.
XX      Disclosure; Fig 1, 121pp; English.
XX      Entire sequence from T4-encoding plasmid p170-2. It is almost identical
CC      to the sequence published by Madden et al.(1985) with the exception of
CC      three codon changes. At T4 amino acid residue 3, (posn. 403 of entire
CC      sequence) Lys is encoded in stead of Asn. At posn. 64, (posn. 464) Arg
CC      replaces Trp and at posn. 231, (posn. 631) Ser replaces Phe. Soluble T4
CC      can be produced by truncating the CDS to remove the transmembrane and
CC      cytoplasmic domains. The soluble forms may be modified to increase their
CC      immunogenicity by addition of an adjuvant such as incomplete Freund's
CC      adjuvant. The T4 interferes with HIV/T4 interaction and elicits anti-
CC      soluble T4 antibody production. See also AAQ05608. (Updated on 31-OCT-
CC      2002 to add missing OS field.)
XX      Sequence 2458 AA;
SQ
Query Match      86.5%; Score 2010; DB 2; Length 2458;
Best Local Similarity 88.8%; Pred. No. 3.8e-128;
Matches 403; Conservative 2; Mismatches 19; Indels 30; Gaps 2;
QY      1  MNRGVPFRHLVLVQLALPAAATGKNKVLGKKGDVLELTCTASQKKSIOFHMKNSNOIK 60
DB      400  MNRGVPFRHLVLVQLALPAAATGKNKVLGKKGDVLELTCTASQKKSIOFHMKNSNOIK 459
QY      61  ILNGGSLFTKGPSKLNDRADSRSLRDQGNPPLIKLKIEDSDTYICEVEDQKEEYQL 120
DB      460  ILNGGSLFTKGPSKLNDRADSRSLRDQGNPPLIKLKIEDSDTYICEVEDQKEEYQL 519
QY      121  LVFGLTANSDFHLLOGSILTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB      520  LVFGLTANSDFHLLOGSILTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 579
QY      181  TWTCTVLQOKKVEFKIDIVLAFOKASSIYKKEGEQVEFSPFLAFVTEKLTGSGELMW 240
DB      580  TWTCTVLQOKKVEFKIDIVLAFOKASSIYKKEGEQVEFSPFLAFVTEKLTGSGELMW 639
QY      241  QAERASSSSKSWITFDLKNKEVSVKRVTDPKIQMGKPLHLTLTLPQALPOYAGSGNLTIA 300
DB      640  QAERASSSSKSWITFDLKNKEVSVKRVTDPKIQMGKPLHLTLTLPQALPOYAGSGNLTIA 699
QY      301  LEAKTGKLEHGVNLVMMRAATOLQKNTCEVNGPTSPKMLSLKLEKKAUKVKREKPPVW 360
DB      700  LEAKTGKLEHGVNLVMMRAATOLQKNTCEVNGPTSPKMLSLKLEKKAUKVKREKPPVW 759
QY      361  LNPEAGMOCCLSDSGQVLLLESNIKVLPTWSTPVHPRASALPAPPTGSLPDPTASALP 420
DB      760  LNPEAGMOCCLSDSGQVLLLESNIKVLPTWSTPVHPRASALPAPPTGSLPDPTASALP 795
QY      421  DPPAASALPALAVISFLIGLGLGV-ACVLAARTR 453
DB      796  -----MALIVLGVAIGLLIFIGIGIFVCVRCRHR 824
RESULT 60
AAQ04031
ID      AAR04031 standard; protein; 2458 AA.
XX      AAR04031;
AC      AAR04031;
XX      25-MAR-2003 (revised)
DT      31-OCT-2002 (revised)
DT      29-MAY-1990 (first entry)
XX      Full length T4 encoded by plasmid p170-2.

```

```

XX      Soluble T4; p170-2; anti-retroviral agent; AIDS; ARC; HIV; AZT.
KM      Synthetic.
XX      Key
XX      Location/Qualifiers
OS      Key
XX      FT      Misc-difference 423
XX      FT      Misc-difference 425 /note= "Trp of Madden et al replaced by Arg"
XX      FT      Misc-difference 425 /note= "Asp of Madden et al replaced by Lys"
XX      FT      Misc-difference 653 /note= "Phe of Madden et al replaced by Ser"
XX      PN      MO8911860-A.
XX      PD      14-DEC-1989.
XX      PF      08-JUN-1989; 89WO-US002453.
XX      PR      10-JUN-1988; 88US-00204645.
XX      PR      20-APR-1989; 89US-00341080.
XX      PA      (BIOJ ) BIOGEN NV INC.
XX      PA      (GENO ) GEN HOSPITAL CORP.
XX      PA      (BIOJ ) BIOGEN INC.
XX      PA      (BIOJ ) BIOGEN INC.
XX      PI      Fisher RA, Schooley RT, Hirsch MS, Johnson VA, Walker BD;
XX      DR      MPI: 1990-007302/01.
XX      DR      N-PSDB; AAQ03005.
XX      PT      Combinations of soluble T4 protein and anti-retroviral agent - having
XX      PT      synergistic activity in treatment and prevention of AIDS, arc and HIV
XX      PS      infection.
XX      PS      Disclosure; Fig 1, 100pp; English.
XX      CC      The sequence differs from that determined by PJ Madden et al., (Cell, 42
XX      CC      pp. 93-104 (1985)) in three places due to three nucleotide substitutions.
XX      CC      The Asp reported at position 3 by Madden et al. was the result of a
XX      CC      sequencing error [DR Littman et al. Cell, 55, p.541 (1988)]. X = stop
XX      CC      codon. The sequence was deduced from the cDNA insert of p170-2. Soluble
XX      CC      T4 constructs may be produced by truncating this sequence to give
XX      CC      fragments from position 400 to 799, removing the transmembrane and
XX      CC      intracytoplasmic domains whilst retaining the extracellular region
XX      CC      responsible for HIV binding. The sol. T4 is combined with an anti-viral
XX      CC      agent such as AZT. (See also AAQ03006. (Updated on 31-OCT-2002 to add
XX      CC      missing OS field.) (Updated on 25-MAR-2003 to correct PA field.)
XX      SQ      Sequence 2458 AA;
Query Match      86.5%; Score 2010; DB 2; Length 2458;
Best Local Similarity 88.8%; Pred. No. 3.8e-128;
Matches 403; Conservative 2; Mismatches 19; Indels 30; Gaps 2;
QY      1  MNRGVPFRHLVLVQLALPAAATGKNKVLGKKGDVLELTCTASQKKSIOFHMKNSNOIK 60
DB      400  MNRGVPFRHLVLVQLALPAAATGKNKVLGKKGDVLELTCTASQKKSIOFHMKNSNOIK 459
QY      61  ILNGGSLFTKGPSKLNDRADSRSLRDQGNPPLIKLKIEDSDTYICEVEDQKEEYQL 120
DB      460  ILNGGSLFTKGPSKLNDRADSRSLRDQGNPPLIKLKIEDSDTYICEVEDQKEEYQL 519
QY      121  LVFGLTANSDFHLLOGSILTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB      520  LVFGLTANSDFHLLOGSILTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 579
QY      181  TWTCTVLQOKKVEFKIDIVLAFOKASSIYKKEGEQVEFSPFLAFVTEKLTGSGELMW 240
DB      580  TWTCTVLQOKKVEFKIDIVLAFOKASSIYKKEGEQVEFSPFLAFVTEKLTGSGELMW 639
QY      241  QAERASSSSKSWITFDLKNKEVSVKRVTDPKIQMGKPLHLTLTLPQALPOYAGSGNLTIA 300

```

```

Db      640 QAEKSSKSMITSDPKKKEVSVKRVTDPKLQWCKKLPHLTLPOALPOYAGSGNLTLA 699
Qy      301 LEAKTGKLGHOENVLVVWRATQLOKNLTCEVWGPTSPKMLSLKLENKAKVSKKEKPVWV 360
Db      700 LEAKTGKLGHOENVLVVWRATQLOKNLTCEVWGPTSPKMLSLKLENKAKVSKKEKAVWV 759
Qy      361 LNPEAGMWQCLLSDSGVLLSESNIKVLPWTWSTPVHPRASALPAPPTGSALPDQTASALP 420
Db      760 LNPEAGMWQCLLSDSGVLLSESNIKVLPWTWSTPVQ----- 795
Qy      421 DPPASALPALAVISFLIGLGV-ACVLARTR 453
Db      796 -----MALIVGGVAGLLFLIGLGFVCRCRHR 824

RESULT 61
AAR11285
ID      AAR11285 standard; protein; 458 AA.
XX
AC      AAR11285;
XX
DT      25-MAR-2003 (revised)
DT      09-JAN-2003 (revised)
DT      29-APR-1991 (first entry)
XX
DE      gp120 binding protein.
XX
KM      Human, CD4; AIDS; HIV1; SIV; gp120.
XX
OS      Unidentified.
XX
FH      Key Location/Qualifiers
FT      Misc-difference 59 /label= Thr or Ile
FT      Misc-difference 93 /label= Thr or Pro
XX
PN      EP414178-A.
XX
PD      27-FEB-1991.
XX
PE      18-AUG-1990; 90EP-00115877.
XX
PR      23-AUG-1989; 89US-00397782.
XX
PA      (GEHO ) GEN HOSPITAL CORP.
XX
PI      Seed B, Camerini D;
XX
PN      WPI; 1991-059419/09.
XX
DR      N-PSDB; AAQ10887.
XX
FT      New non-human primate and human CD4 or gp120 molecules - used to treat
FT      HIV or SIV and immunoglobulin and gp120 binding molecules from new fusion
FT      proteins.
XX
PS      Claim 17; Page 57; 87pp; English.
XX
CC      The fragment from residues 1-134 is also independently claimed. The sub-
CC      fragment (and the complete polypeptide) can bind to HIV gp120. See also
CC      AAQ10885-6, AAQ10888. (Updated on 09-JAN-2003 to add missing OS field.)
CC      (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ      Sequence 458 AA;

Query Match      86.4%; Score 2008; DB 2; Length 458;
Best Local Similarity 88.3%; Pred. No. 6-7e-129;
Matches 401; Conservative 3; Mismatches 20; Indels 30; Gaps 2;
Qy      1 MNRGVFPHLLVLTALPAPATQGNKVVLGKGDPTVELTCTASQKSIQFMWKSNOIK 60
Db      1 MNRGVFPHLLVLTALPAPATQGNKVVLGKGDPTVELTCTASQKSIQFMWKSNOIK 60

```

```

Qy      61 ILGNQGSFLTKGPSKLNDRADSRSLMDGNPFLIKNLKIEDSDTYICEVEDOKEEVQL 120
Db      61 ILGNQGSFLTKGPSKLNDRADSRSLMDGNPFLIKNLKIEDSDTYICEVEDOKEEVQL 120
Qy      121 LVPEGLTANSDTHLLDQGSLLTLESPPGSSPVQCRSPRGKNIQGGKTLASVQLELQDSG 180
Db      121 LVPEGLTANSDTHLLDQGSLLTLESPPGSSPVQCRSPRGKNIQGGKTLASVQLELQDSG 180
Qy      181 TWCTTVQNKQKVEFKDIIYVLAFOKASSIVYKKEGQVFPFLATVTEKLTSSGSLMW 240
Db      181 TWCTTVQNKQKVEFKDIIYVLAFOKASSIVYKKEGQVFPFLATVTEKLTSSGSLMW 240
Qy      241 QAEKSSKSMITFDLKNKEVSVKRVTDPKLQWCKKLPHLTLPOALPOYAGSGNLTLA 300
Db      241 QAEKSSKSMITFDLKNKEVSVKRVTDPKLQWCKKLPHLTLPOALPOYAGSGNLTLA 300
Qy      301 LEAKTGKLGHOENVLVVWRATQLOKNLTCEVWGPTSPKMLSLKLENKAKVSKKEKPVWV 360
Db      301 LEAKTGKLGHOENVLVVWRATQLOKNLTCEVWGPTSPKMLSLKLENKAKVSKKEKAVWV 360
Qy      361 LNPEAGMWQCLLSDSGVLLSESNIKVLPWTWSTPVHPRASALPAPPTGSALPDQTASALP 420
Db      361 LNPEAGMWQCLLSDSGVLLSESNIKVLPWTWSTPVQ----- 396
Qy      421 DPPASALPALAVISFLIGLGV-ACVLARTR 453
Db      397 -----MALIVGGVAGLLFLIGLGFVCRCRHR 425

RESULT 62
AAR10988
ID      AAR10988 standard; protein; 458 AA.
XX
AC      AAR10988;
XX
DT      25-MAR-2003 (revised)
DT      29-APR-1991 (first entry)
XX
DE      Chimpanzee CD4 protein.
XX
KM      Chimpanzee; CD4; AIDS; HIV1; SIV.
XX
OS      Pan troglodytes.
XX
FH      Key Location/Qualifiers
FT      Protein 26 /label= mature CD4
XX
PN      EP414178-A.
XX
PD      27-FEB-1991.
XX
PE      18-AUG-1990; 90EP-00115877.
XX
PR      23-AUG-1989; 89US-00397782.
XX
PA      (GEHO ) GEN HOSPITAL CORP.
XX
PI      Seed B, Camerini D;
XX
PN      WPI; 1991-059419/09.
XX
DR      N-PSDB; AAQ10886.
XX
FT      New non-human primate and human CD4 or gp120 molecules - used to treat
FT      HIV or SIV and immunoglobulin and gp120 binding molecules from new fusion
FT      proteins.
XX
PS      Claim 4; Page 45; 87pp; English.
XX
CC      The CD4 protein or HIV gp120-binding fragments of it are used to detect
CC      and treat HIV and SIV infection. Animals which can be treated include
CC      humans, baboons, orang-utans, chimpanzees, gorillas and rhesus monkeys.

```



CC The chimpanzee CD4 is 99 per cent homologous to its human counterpart,  
CC possessing 5 amino acid substitutions in the 433 residue predicted mature  
CC polypeptide. See also AAQ10885, AAQ10887-8. (Updated on 25-MAR-2003 to  
CC correct PA field.)

XX Sequence 458 AA;

Query Match 86.4%; Score 2008; DB 2; Length 458;  
Best Local Similarity 88.3%; Pred. No. 6,7e-129;  
Matches 401; Conservative 2; Mismatches 21; Indels 30; Gaps 2;

QY 1 NMRGVFPHLLVLTQALPAATQGNKVLGKKGDVETLTASQKKSIOFHMKNSNOIK 60  
DB 1 NMRGVFPHLLVLTQALPAATQGNKVLGKKGDVETLTASQKKSIOFHMKNSNOIK 60  
QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFTLLIKNKIKEDSDTYICEVDQKEEYQL 120  
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFTLLIKNKIKEDSDTYICEVDQKEEYQL 120  
QY 121 LVFGLTANSPTHLLQGSLLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
DB 121 LVFGLTANSPTHLLQGSLLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
QY 181 TWCTVLTQNKQKVEFKIDIVLAFOKASSIYKKEGEQVEFSFPLAFTVEKLTGSGELMW 240  
DB 181 TWCTVLTQNKQKVEFKIDIVLAFOKASSIYKKEGEQVEFSFPLAFTVEKLTGSGELMW 240  
QY 241 QAEKRASSKSWITPDLKKEVSVKRVTDPKIQMGKPLHLTLTPQALPOYAGSNTLTA 300  
DB 241 QAEKRASSKSWITPDLKKEVSVKRVTDPKIQMGKPLHLTLTPQALPOYAGSNTLTA 300  
QY 301 LEAKTGKLGHOEVNLVVMRATOLQKNTCEVWGPTSPKMLSLKLENKEAKVSKREKPYWV 360  
DB 301 LEAKTGKLGHOEVNLVVMRATOLQKNTCEVWGPTSPKMLSLKLENKEAKVSKREKPYWV 360  
QY 361 LNPEAGMOCCLSDSGVLTLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPTASALP 420  
DB 361 LNPEAGMOCCLSDSGVLTLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPTASALP 420  
QY 421 DPPAASALPALAVISFLIGLGLGV-ACVLAATR 453  
DB 397 -----MALIVLGVAGLLFTIGLGFVCRCRHR 425

# RESULT 63

AAQ04910  
ID AAQ04910 standard; protein; 458 AA.

XX AAR04910;  
AC 31-OCT-2002 (revised)  
DT 02-OCT-1990 (first entry)  
XX T4 protein as encoded by p170.2.  
DE T4 protein; immunotoxin; Pseudomonas endotoxin A; AIDS; HIV; ARC;  
KM angiotensin; fusion protein.  
OS Synthetic.  
XX Key Location/Qualifiers  
FH Peptide 1..26  
FT Protein /label= signal peptide  
FT Protein /label= T4 protein  
FT M18c-difference replace(87..6)  
FT M18c-difference /note= "differs from Madden et al"  
FT M18c-difference replace(254..F)  
FT /note= "differs from Madden et al"  
XX WO9004414-A.  
XX 03-MAY-1990.  
PD

XX 18-OCT-1988; 88US-00259355.  
XX 18-OCT-1988; 88US-00259355.  
XX 18-OCT-1988; 88US-00259355.  
XX (BIOJ ) BIOGEN INC.  
XX Meade HM, Lobb RR, Gates LL, Winkler G;  
XX WPI, 1990-163876/21.  
XX N-PSDB; AAQ04555.

PT New immunotoxin contg. soluble T4 protein components and toxin - esp.  
PT Pseudomonas endotoxin A, for treating or controlling AIDS and related  
PT conditions, and new DNA sequences.

PS Disclosure; Page ?; -pp; English.

CC The T4 protein encoded by p170.2 is almost identical to that reported by  
CC P.J. Madden et al [Cell, 42, pp 93-104 (1985)]. The Madden sequence was  
CC revised in 1988 to correct a DNA sequencing error at AA 3 (corrected from  
CC Asp to Lys; see M12807 in Genbank). The DNA may be truncated (to remove  
CC transmembrane and intracellular regions) and/or modified by SDM, pref. so  
CC the prod. extends from AAs 3-183 of the mature protein. This DNA can then  
CC be ligated to a toxin DNA esp. angiotensin, or a fragment of Pseudomonas  
CC exotoxin A contg. the translocation and ADP-ribosylation domains. The  
CC hybrid DNA can then be inserted into an expression vector and used to  
CC produce recombinant fusion protein which is useful for preventing or  
CC treating AIDS, ARC, and HIV infections. The T4 protein is an HIV receptor  
CC which binds to virus or to infected cells carrying the gp120/160 marker  
CC antigen, so provides v. specific targeting with minimal damage to non-  
CC target cells. (Updated on 31-OCT-2002 to add missing OS field.)

XX Sequence 458 AA;

Query Match 86.1%; Score 2002; DB 2; Length 458;  
Best Local Similarity 88.3%; Pred. No. 1.7e-128;  
Matches 401; Conservative 4; Mismatches 19; Indels 30; Gaps 2;

QY 1 NMRGVFPHLLVLTQALPAATQGNKVLGKKGDVETLTASQKKSIOFHMKNSNOIK 60  
DB 1 NMRGVFPHLLVLTQALPAATQGNKVLGKKGDVETLTASQKKSIOFHMKNSNOIK 60  
QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFTLLIKNKIKEDSDTYICEVDQKEEYQL 120  
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFTLLIKNKIKEDSDTYICEVDQKEEYQL 120  
QY 121 LVFGLTANSPTHLLQGSLLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
DB 121 LVFGLTANSPTHLLQGSLLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
QY 181 TWCTVLTQNKQKVEFKIDIVLAFOKASSIYKKEGEQVEFSFPLAFTVEKLTGSGELMW 240  
DB 181 TWCTVLTQNKQKVEFKIDIVLAFOKASSIYKKEGEQVEFSFPLAFTVEKLTGSGELMW 240  
QY 241 QAEKRASSKSWITPDLKKEVSVKRVTDPKIQMGKPLHLTLTPQALPOYAGSNTLTA 300  
DB 241 QAEKRASSKSWITPDLKKEVSVKRVTDPKIQMGKPLHLTLTPQALPOYAGSNTLTA 300  
QY 301 LEAKTGKLGHOEVNLVVMRATOLQKNTCEVWGPTSPKMLSLKLENKEAKVSKREKPYWV 360  
DB 301 LEAKTGKLGHOEVNLVVMRATOLQKNTCEVWGPTSPKMLSLKLENKEAKVSKREKPYWV 360  
QY 361 LNPEAGMOCCLSDSGVLTLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPTASALP 420  
DB 361 LNPEAGMOCCLSDSGVLTLESNIKVLPTWSTPVHPRASALPAPPTGSALPDPTASALP 420  
QY 421 DPPAASALPALAVISFLIGLGLGV-ACVLAATR 453  
DB 397 -----MALIVLGVAGLLFTIGLGFVCRCRHR 425

# RESULT 64

AAB07768  
 ID AAB07768 standard; protein; 394 AA.  
 XX  
 AC AAB07768;  
 XX  
 DT 07-NOV-2000 (first entry)  
 XX  
 DE The soluble extracellular domain of the T4 glycoprotein.  
 XX  
 KM Human; T4 glycoprotein; human immunodeficiency virus; HIV;  
 KM envelope glycoprotein; AIDS; virus binding.  
 XX  
 OS Homo sapiens.  
 OS  
 PN US6093539-A.  
 PD 25-JUL-2000.  
 XX  
 PF 06-JUN-1995; 95US-00466368.  
 XX  
 PR 21-AUG-1986; 86US-00898587.  
 PR 11-JUN-1991; 91US-00713564.  
 PR 06-JUL-1992; 92US-00909021.  
 PR 12-DEC-1994; 94US-00354452.  
 XX  
 PA (UYCO ) UNIV COLUMBIA NEW YORK.  
 XX  
 PI Maddon PJ, Chess L, Axel R, Weiss R, McDougal JS, Littman DR;  
 DR WPI; 2000-505203/45.  
 DR N-PSDB; AAA59351.  
 XX  
 PT New isolated nucleic acid encoding a human T cell surface protein and the  
 PT soluble surface T4 glycoprotein that it encodes, useful as prophylaxis  
 PT for creating a subject infected with human acquired immune deficiency  
 PT syndrome virus.  
 XX  
 PS Disclosure; Col 11-14; 69pp; English.  
 XX  
 CC The present sequence represents an aqueous-soluble polypeptide comprising  
 CC a portion of a human T4 glycoprotein. The portion specifically forms a  
 CC complex with a human immunodeficiency virus (HIV) envelope glycoprotein.  
 CC The DNA is useful for producing the soluble surface T4 glycoprotein. The  
 CC soluble surface T4 glycoprotein is useful as a therapeutic agent, i.e. as  
 CC prophylaxis for treating a subject infected with an HIV virus. Thus, the  
 CC soluble T4 glycoprotein is useful for treating human AIDS. The soluble T4  
 CC glycoprotein is also useful in diagnostic or screening assays, e.g. for  
 CC screening inhibitors of virus binding, or for detecting and quantitating  
 CC T4, T4+ cells and antibodies to T4, which are of diagnostic value for  
 CC AIDS  
 CC  
 CC Sequence 394 AA;  
 SQ  
 Query Match 86.1%; Score 2001; DB 3; Length 394;  
 Best Local Similarity 98.7%; Pred. No. 1.7e-128;  
 Matches 389; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 MNRGVPFRHLVQLALPPATQGNKVLGKKGDTVELTCTASQKSIQFHMNSNQIK 60  
 DB 1 MNRGVPFRHLVQLALPPATQGNKVLGKKGDTVELTCTASQKSIQFHMNSNQIK 60  
 QY 1 ILNGSGFLTKGSPKLTNDRAISRSLMDQGNFPLIKLKLTEDSDTYICEVEDQKEVQL 120  
 DB 61 ILNGSGFLTKGSPKLTNDRAISRSLMDQGNFPLIKLKLTEDSDTYICEVEDQKEVQL 120  
 QY 121 LVFGLTANSDTHLLQGSGSLTLTLESPGSSPSVQCRSPRKNIOGKTLVSQLELQDSG 180  
 DB 121 LVFGLTANSDTHLLQGSGSLTLTLESPGSSPSVQCRSPRKNIOGKTLVSQLELQDSG 180  
 QY 181 TWTCIVLONOKKVEKIDIVVLAPOKASIVYKKGGEVSSFPALFVETLTSSELGMW 240  
 DB 181 TWTCIVLONOKKVEKIDIVVLAPOKASIVYKKGGEVSSFPALFVETLTSSELGMW 240

QY 241 QAERASSSKSMITFDLKNKEVSVKRVTPQDKQMGKLPALTLTPQALPOYAGSGNLTIA 300  
 DB 241 QAERASSSKSMITFDLKNKEVSVKRVTPQDKQMGKLPALTLTPQALPOYAGSGNLTIA 300  
 QY 301 LEAKTGKHQEVNLVVMRATQKNTLCEVWGPTSPKMLSLKLENKARVSKREKPVW 360  
 DB 301 LEAKTGKHQEVNLVVMRATQKNTLCEVWGPTSPKMLSLKLENKARVSKREKPVW 360  
 QY 361 LNPEAGMWOCLLSDSGVLLSESNIKVLPWTSTPV 394  
 DB 361 LNPEAGMWOCLLSDSGVLLSESNIKVLPWTSTPV 394  
 RESULT 65  
 ID AAR20151 standard; protein; 399 AA.  
 XX  
 AC AAR20151;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 31-MAR-1992 (first entry)  
 XX  
 DE Chimpanzee sol. CD4 encoded by pSQ200.  
 XX  
 KM Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;  
 KM acquired immune deficiency syndrome; AIDS related complex;  
 KM T helper lymphocytes.  
 XX  
 OS Pan troglodytes.  
 OS  
 FH Key Location/Qualifiers  
 FT Peptide 1..25  
 FT /label= signal\_sequence  
 XX  
 PN W09118618-A.  
 PD 12-DEC-1991.  
 XX  
 PF 25-MAY-1990; 90US-00529186.  
 XX  
 PR 25-MAY-1990; 90US-00529186.  
 XX  
 PA (BIOJ ) BIOGEN INC.  
 PI Fisher RA, Hession C, Burkiy LC;  
 DR WPI; 1992-007200/01.  
 DR N-PSDB; AAQ20326.  
 XX  
 PT New immuno-therapeutic human CD4 variants and derivs. - elicit AB  
 PT production to HIV gp 120, useful in treating, preventing and diagnosing  
 PT AIDS, ARC and HIV infections.  
 XX  
 PS Claim 15; Fig 21; 179pp; English.  
 XX  
 CC The sequence was deduced from the DNA sequence of clone pSQ200 which was  
 CC obtd. by cloning using a reverse transcriptase/PCR amplification  
 CC procedure with a non-specific oligo dt to prime first strand synthesis.  
 CC The PCR fragments were ligated into sequencing plasmid pNN08 (see  
 CC AAQ20326) The DNA sequence can be used to express recombinant soluble CD4  
 CC and analogues for use in diagnosis and treatment of diseases caused by  
 CC infective agents whose primary targets are T4+ lymphocytes. See also  
 CC AAR20148-R20155 and AAR21078. (updated on 25-MAR-2003 to correct PA  
 CC field.)  
 CC  
 CC Sequence 399 AA;  
 SQ  
 Query Match 86.1%; Score 2001; DB 2; Length 399;  
 Best Local Similarity 98.0%; Pred. No. 1.7e-128;  
 Matches 390; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 MNRGVPFRHLVQLALPPATQGNKVLGKKGDTVELTCTASQKSIQFHMNSNQIK 60

```

Db      1 MNRGVPFRHLILVLQALLPAATGKVVLGKKGDVTELTCTASQKKSIOFHMKNNSQTK 60
Qy      61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVDDQKEEVOL 120
Db      61 ILGNQGSFLTKGPSKLNDRVDSRSLMDQGNFPLIIKNLKIEDSDTYICEVDDQKEEVOL 120
Qy      121 LVFGLTANSDFTHLQGSILTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQDSG 180
Db      121 LVFGLTANSDFTHLQGSILTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQDSG 180
Qy      181 TWTCTVLQNOKKVEFKIDIVLAFQKASSIYKKEGEQVEFSFPLAFTVEKLTGSGELMW 240
Db      181 TWTCTVLQNOKKVEFKIDIVLAFQKASSIYKKEGEQVEFSFPLAFTVEKLTGSGELMW 240
Qy      241 QAERASSSKSWITTFDLKNKEVSVKRVTQDPKLOMGKKLPLHLTLPOALPOYAGSGNLTLA 300
Db      241 QAERASSSKSWITTFDLKNKEVSVKRVTQDPKLOMGKKLPLHLTLPOALPOYAGSGNLTLA 300
Qy      301 LEAKTGKLEHOEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPYAV 360
Db      301 LEAKTGKLEHOEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPYAV 360
Qy      361 LNPEAGMWQCLSDSGQVLLSNIKVLPWTWSTPVPMA 398
Db      361 LNPEAGMWQCLSDSGQVLLSNIKVLPWTWSTPVPMA 398

RESULT 66
AAR20150
ID      AAR20150 standard; protein; 400 AA.
XX
XX      AAR20150;
AC
XX
XX      25-MAR-2003 (revised)
DT
XX      31-MAR-1992 (first entry)
DE      Chimpanzee sol. CD4 sequence from psQ205.
XX
XX      Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;
KM      acquired immune deficiency syndrome; AIDS related complex;
KW      T helper lymphocytes.
XX
XX      Pan troglodytes.
OS
XX
XX      Key      Location/Qualifiers
FT      Peptide      1..25
FT      /label= signal_sequence
XX
XX      MO9118618-A.
XX
XX      12-DEC-1991.
PD
XX
XX      25-MAY-1990; 90US-00529186.
PF
XX
XX      25-MAY-1990; 90US-00529186.
PR
XX
XX      (BIOG ) BIOGEN INC.
XX
XX      Fisher RA, Hession C, Burkly LC;
XX
XX      WPI; 1992-007200/01.
XX      N-PSDB; AAQ20325.
DR
XX
XX      New immuno-therapeutic human CD4 variants and derivs. - elicit AB
PT      production to HIV gp.120, useful in treating, preventing and diagnosing
PT      AIDS, ARC and HIV infections.
XX
XX      Claim 15; Fig 20; 179pp; English.
XX
XX      The sequence was deduced from the DNA sequence of clone psQ205 which was
CC      obtd. by cloning using a reverse transcriptase/PCR amplification
CC      procedure. The DNA sequence can be used to express recombinant soluble
CC      CD4 and analogues for use in diagnosis and treatment of diseases caused

```

```

CC      by infective agents whose primary targets are T4+ lymphocytes. See also
CC      AAR20148-R20155 and AAR21078. (Updated on 25-MAR-2003 to correct PA
CC      field.)
XX
XX      SQ      Sequence 400 AA;
Qy      Query Match      85.8%; Score 1996; DB 2; Length 400;
Db      Best Local Similarity 97.7%; Pred. No. 3,8e-128;
Matches 389; Conservative 1; Mismatches 8; Indels 0; Gaps 0;
Qy      1 MNRGVPFRHLILVLQALLPAATGKVVLGKKGDVTELTCTASQKKSIOFHMKNNSQTK 60
Db      1 MNRGVPFRHLILVLQALLPAATGKVVLGKKGDVTELTCTASQKKSIOFHMKNNSQTK 60
Qy      61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVDDQKEEVOL 120
Db      61 ILGNQGSFLTKGPSKLNDRVDSRSLMDQGNFPLIIKNLKIEDSDTYICEVDDQKEEVOL 120
Qy      121 LVFGLTANSDFTHLQGSILTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQDSG 180
Db      121 LVFGLTANSDFTHLQGSILTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQDSG 180
Qy      181 TWTCTVLQNOKKVEFKIDIVLAFQKASSIYKKEGEQVEFSFPLAFTVEKLTGSGELMW 240
Db      181 TWTCTVLQNOKKVEFKIDIVLAFQKASSIYKKEGEQVEFSFPLAFTVEKLTGSGELMW 240
Qy      241 QAERASSSKSWITTFDLKNKEVSVKRVTQDPKLOMGKKLPLHLTLPOALPOYAGSGNLTLA 300
Db      241 QAERASSSKSWITTFDLKNKEVSVKRVTQDPKLOMGKKLPLHLTLPOALPOYAGSGNLTLA 300
Qy      301 LEAKTGKLEHOEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPYAV 360
Db      301 LEAKTGKLEHOEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPYAV 360
Qy      361 LNPEAGMWQCLSDSGQVLLSNIKVLPWTWSTPVPMA 398
Db      361 LNPEAGMWQCLSDSGQVLLSNIKVLPWTWSTPVPMA 398

RESULT 67
AAP93010
ID      AAP93010 standard; protein; 399 AA.
XX
XX      AAP93010;
AC
XX
XX      25-MAR-2003 (revised)
DT
XX      03-AUG-1992 (first entry)
DE      Genetic construct which encodes CD4 linked to human IGM at the Mst2 site
DE      upstream of the CHI region (fusion protein CD44mu).
XX
XX      Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy;
KW      diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.
XX
XX      Homo sapiens.
OS
XX
XX      EP325262-A.
XX
XX      26-JUL-1989.
PD
XX
XX      20-JAN-1989; 89EP-00100913.
PF
XX
XX      22-JAN-1988; 88US-00147351.
PR
XX
XX      (GEHO ) GEN HOSPITAL CORP.
XX
XX      Seed B;
XX
XX      WPI; 1989-214472/30.
XX      N-PSDB; AAN90358.
DR
XX
XX      Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV
PT      infections or detecting HIV or SIV in sample.

```

XX Example: Table 3, Page 34-40; 68pp; English.

CC The fusion protein genes of the invention pref. comprises cDNA sequences  
CC which encode CD4 or a fragment which binds gp120 ligated to an expression  
CC plasmid which encodes an antibody in which the variable region of the  
CC gene has been deleted (see WO87-02671). The CD4 portion of the fusion  
CC protein may comprise the complete CD4 sequence, the 370 AA extracellular  
CC region and the membrane spanning domain, or the extracellular region. The  
CC Ig heavy chain is pref. from IGM, IgG1 or IgG3. The following are  
CC specifically claimed: fusion proteins CD4H1ambdai, CD4Mmu, CD4Fmu,  
CC CD4E1ambdai, and CD4Ymu (No. 67608), PCd4plambda (No. 67609) and  
CC PCd4E1ambdai (No. 67610). (Updated on 25-MAR-2003 to correct PA field.)

XX Sequence 399 AA;

Query Match 85.6%; Score 1990; DB 1; Length 399;  
Best Local Similarity 98.5%; Pred. No. 9,6e-128;  
Matches 389; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```
QY 1 MNRGVPFRHLLLVQLALLPAATQGNKRVLGKKGDTVELTCTASQKSIQFHMKNNSNQIK 60
DB 1 MNRGVPFRHLLLVQLALLPAATQGNKRVLGKKGDTVELTCTASQKSIQFHMKNNSNQIK 60
QY 61 ILNGGSEFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVOL 120
DB 61 ILNGGSEFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGITANSPTHLLOGSLTLTLSPSPSSPSVQCRSPRGNIOGKTLVSQLELQDSG 180
DB 121 LVFGITANSPTHLLOGSLTLTLSPSPSSPSVQCRSPRGNIOGKTLVSQLELQDSG 180
QY 121 LVFGITANSPTHLLOGSLTLTLSPSPSSPSVQCRSPRGNIOGKTLVSQLELQDSG 180
DB 121 LVFGITANSPTHLLOGSLTLTLSPSPSSPSVQCRSPRGNIOGKTLVSQLELQDSG 180
QY 181 TWTCVTLONOKKVEFKIDIVLAFOKASSIYKKEGEVSEFPLAFVEKLTGSGELMW 240
DB 181 TWTCVTLONOKKVEFKIDIVLAFOKASSIYKKEGEVSEFPLAFVEKLTGSGELMW 240
QY 181 TWTCVTLONOKKVEFKIDIVLAFOKASSIYKKEGEVSEFPLAFVEKLTGSGELMW 240
DB 181 TWTCVTLONOKKVEFKIDIVLAFOKASSIYKKEGEVSEFPLAFVEKLTGSGELMW 240
QY 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKULPLHLTPQALPOYAGSGNLTLA 300
DB 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKULPLHLTPQALPOYAGSGNLTLA 300
QY 301 LEAKTGKLEHVEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKAQVSKREKPVW 360
DB 301 LEAKTGKLEHVEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKAQVSKREKPVW 360
QY 361 LNPEAGMOCCLSDSGVLLSESNIKVLPWTSTPVH 395
DB 361 LNPEAGMOCCLSDSGVLLSESNIKVLPWTSTPVH 395
```

RESULT 68  
AAB19511  
ID AAB19511 standard; protein; 616 AA.

XX AAB19511;  
AC AAB19511;  
XX 09-JAN-2001 (first entry)  
XX  
DE CD4-IgG1 fusion protein CH4Bgammal.  
XX  
XX CD4; IgG1; human; CD4Bgammal; fusion protein; immunoglobulin; HIV; SIV;  
KW gp120; therapy; diagnosis.  
XX  
XX Homo sapiens.  
XX  
FH Key  
FT Protein  
FT 1. .395 Location/Qualifiers  
FT /note= "CD4 extracellular region"  
FT 400. .616  
FT /note= "IgG1 heavy chain"  
XX  
XX US6117656-A.  
XX  
XX 12-SEP-2000.

XX 07-JUN-1995; 95US-00479353.

XX 22-JAN-1988; 88US-00147351.

XX 23-JAN-1989; 89US-00299596.

XX 09-JUN-1992; 92US-00896781.

XX 12-APR-1993; 93US-00057952.

XX 04-FEB-1994; 94US-00191708.

XX (GENO ) GEN HOSPITAL CORP.

XX Seed B;

XX WPI; 2000-58658/55.

XX N-PSDB; AAA50664.

XX CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or SIV.

XX Example 1; Col 59-70; 39pp; English.

CC The present sequence is that of fusion protein CD4Bgammal comprising the  
CC extracellular portion of CD4, which binds to HIV gp120, linked at its C-  
CC terminus to the human IgG1 heavy chain. To obtain the fusion protein, DNA  
CC encoding CD4 was linked to IgG1 DNA at the BamI site downstream of the  
CC hinge region (see AAA50664). Fusion protein CD4Bgammal and a nucleic acid  
CC encoding it are claimed. Also claimed are a vector comprising the nucleic  
CC acid, and a method of producing the fusion protein in secreted form using  
CC a transformed host cell. The fusion protein may further comprise a  
CC therapeutic agent, radiolabel or NMR imaging agent. The fusion protein  
CC can be administered to an animal (including humans) for treatment of HIV  
CC or SIV infection, and can also be used in assays for HIV or SIV, imaging  
CC and tissue stains. IgG1 fusion proteins such as CD4Bgammal provide both  
CC complement-mediated and cell-mediated immunity

XX Sequence 616 AA;

Query Match 84.9%; Score 1974; DB 3; Length 616;  
Best Local Similarity 94.9%; Pred. No. 2e-126;  
Matches 392; Conservative 1; Mismatches 18; Indels 2; Gaps 1;

```
QY 1 MNRGVPFRHLLLVQLALLPAATQGNKRVLGKKGDTVELTCTASQKSIQFHMKNNSNQIK 60
DB 1 MNRGVPFRHLLLVQLALLPAATQGNKRVLGKKGDTVELTCTASQKSIQFHMKNNSNQIK 60
QY 61 ILNGGSEFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVOL 120
DB 61 ILNGGSEFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVOL 120
QY 61 ILNGGSEFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVOL 120
DB 61 ILNGGSEFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGITANSPTHLLOGSLTLTLSPSPSSPSVQCRSPRGNIOGKTLVSQLELQDSG 180
DB 121 LVFGITANSPTHLLOGSLTLTLSPSPSSPSVQCRSPRGNIOGKTLVSQLELQDSG 180
QY 121 LVFGITANSPTHLLOGSLTLTLSPSPSSPSVQCRSPRGNIOGKTLVSQLELQDSG 180
DB 121 LVFGITANSPTHLLOGSLTLTLSPSPSSPSVQCRSPRGNIOGKTLVSQLELQDSG 180
QY 181 TWTCVTLONOKKVEFKIDIVLAFOKASSIYKKEGEVSEFPLAFVEKLTGSGELMW 240
DB 181 TWTCVTLONOKKVEFKIDIVLAFOKASSIYKKEGEVSEFPLAFVEKLTGSGELMW 240
QY 181 TWTCVTLONOKKVEFKIDIVLAFOKASSIYKKEGEVSEFPLAFVEKLTGSGELMW 240
DB 181 TWTCVTLONOKKVEFKIDIVLAFOKASSIYKKEGEVSEFPLAFVEKLTGSGELMW 240
QY 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKULPLHLTPQALPOYAGSGNLTLA 300
DB 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKULPLHLTPQALPOYAGSGNLTLA 300
QY 301 LEAKTGKLEHVEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKAQVSKREKPVW 360
DB 301 LEAKTGKLEHVEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKAQVSKREKPVW 360
QY 361 LNPEAGMOCCLSDSGVLLSESNIKVLPWTSTPVH--PRASALPAPPTGALP 411
DB 361 LNPEAGMOCCLSDSGVLLSESNIKVLPWTSTPVHADPEAPBELIGSPVFLFP 413
```

RESULT 69  
AAR20154  
ID AAR20154 standard; protein; 400 AA.

```

XX AC AAR20154;
XX DT 24-OCT-2003 (revised)
XX DT 25-MAR-2003 (revised)
XX DT 31-MAR-1992 (first entry)
XX DE Sol. rhesus-human chimeric CD4 encoded by pBG100.
XX KW Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;
XX KW acquired immune deficiency syndrome; AIDS related complex; monkey;
XX KW T helper lymphocytes.
XX OS Macaca mulatta; (Rhesus).
XX OS Homo sapiens.
XX OS Chimeric.
XX FH Key Location/Qualifiers
XX FH Region 1..131
XX FT /note="rhesus CD4 encoded by pBG146 (AAQ20328)"
XX FT Peptide 1..25
XX FT /label= signal_sequence
XX FT Region 132..400
XX FT /note="human CD4 encoded by pBG391 (US8802940)"
XX PN MO9118618-A.
XX PD 12-DEC-1991.
XX PF 25-MAY-1990; 90US-00529186.
XX PR 25-MAY-1990; 90US-00529186.
XX PA (BIOJ ) BIOGEN INC.
XX PI Fisher RA, Hession C, Burkly LC;
XX DR WPI: 1992-007200/01.
XX DR N-PSDB; AAQ20329.
XX PT New immuno-therapeutic human CD4 variants and derivs. - elicit AB
XX PT production to HIV gp.120, useful in treating, preventing and diagnosing
XX PT AIDS, ARC and HIV infections.
XX PS Claim 14; Fig 13; 179pp; English.
XX CC The sequence is encoded by pBG100, chimeric plasmid prepd. from human and
XX CC rhesus derived CD4 DNA. The plasmid can be used to express recombinant
XX CC sol. chimeric CD4 for use in diagnosis and treatment of diseases caused
XX CC by infective agents whose primary targets are T4+ lymphocytes. See also
XX CC AAR20148-R20155 and AAR21078. (Updated on 25-MAR-2003 to correct PA
XX CC field.) (Updated on 24-OCT-2003 to standardise OS field)
XX SQ Sequence 400 AA;

Query Match 84.1%; Score 1956; DB 2; Length 400;
Best Local Similarity 95.2%; Pred. No. 2e-125;
Matches 379; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

```

```

Db 181 TWTCTVLQNKKEFKIDIVLAFQKASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMW 240
Qy 241 QAERASSKSWITPDLKRKEVSVKRVTPDPLQMGKKLPLHLTLPOALPOYAGSGNLTLA 300
Db 241 QAERASSKSWITPDLKRKEVSVKRVTPDPLQMGKKLPLHLTLPOALPOYAGSGNLTLA 300
Qy 301 LEAKTGKHOENVLVMMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKYSKEKPYAV 360
Db 301 LEAKTGKHOENVLVMMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKYSKEKPYAV 360
Qy 361 LNPEAGMWQCLLSDSGQVLLBSNISKVLPWTGTPVHPRA 398
Db 361 LNPEAGMWQCLLSDSGQVLLBSNISKVLPWTGTPVHPMA 398

RESULT 70
AAR15149
ID AAR15149 standard; protein; 458 AA.
XX AC AAR15149;
XX DT 25-MAR-2003 (revised)
XX DT 24-FEB-1992 (first entry)
XX DE CD4 coordinate system.
XX KW Gelsolin; fusion protein; diagnosis; AIDS.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FH Domain 1..25
XX FT /label= hydrophobic/secretory_signal
XX FT Domain 26..132
XX FT /label= first_Ig-related_domain
XX FT /note= "extracellular"
XX FT Disulfide-bond 41..109
XX FT Domain 133..202
XX FT /label= second_Ig-related_domain
XX FT /note= "extracellular"
XX FT Disulfide-bond 155..184
XX FT Domain 203..318
XX FT /label= third_Ig-related_domain
XX FT /note= "extracellular"
XX FT Domain 319..395
XX FT /label= fourth_Ig-related_domain
XX FT /note= "extracellular"
XX FT Disulfide-bond 328..370
XX FT Domain 396..416
XX FT /label= hydrophobic/transmembrane_sequence
XX FT Domain 417..458
XX FT /label= very_hydrophilic/intracytoplasmic

MO9117170-A.
XX PN
XX PD 14-NOV-1991.
XX PF 04-MAY-1990; 90US-00520368.
XX PR 04-MAY-1990; 90US-00520368.
XX PA (BIOJ ) BIOGEN INC.
XX PI Pepinsky RB, Rosa MD, Stosbel TP;
XX DR WPI: 1991-353711/48.
XX DR N-PSDB; AAQ14931.
XX PT New multi-meric and hetero-multi-meric gelsolin fusion constructs - used
XX PT to treat and diagnose AIDS, ARC and HIV infection.
XX PS Disclosure; Fig 3A-3D; 131pp; English.
XX XX

```

CC The CD4 polypeptides useful in the constructs include all CD4  
 CC polypeptides which bind to or otherwise inhibit gp120 and gp160. These  
 CC include fragments lacking the transmembrane domain. In particular it is  
 CC CD4 1-111; 1-Cys111; 1-Cys180; 1-181, 1-183; 1-187; 1-345 or 1-375 (from  
 CC mature protein). See also AA014931-35 and AAR15151. (Updated on 25-MAR-  
 CC 2003 to correct PA field.)

CC Sequence 458 AA;

Query Match 83.2%; Score 1934; DB 2; Length 458;  
 Best Local Similarity 85.2%; Pred. No. 7,6e-124;  
 Matches 387; Conservative 9; Mismatches 28; Indels 30; Gaps 2;

QY 1 MNRGVFPHLLVLTQALLPAATQGNKVLGKGGDTVELCTASQKSIQFHMKNNSQIK 60  
 DB 1 MNRGVFPHLLVLTQALLPAATQGNKVLGKGGDTVELCTASQKSIQFHMKNNSQIK 60  
 QY 61 ILGNQSFLLTKGPKSLNDRADSRSLMDQGNFPLIIKMLKIEDSDTYICEVEDQKEEYQL 120  
 DB 61 ILGNQSFLLTKGPKSLNDRADSRSLMDQGNFPLIIKMLKIEDSDTYICEVEDQKEEYQL 120  
 QY 121 LVFGLTANSDPHLLQGSGLTLTLESPGSSPSVQCRSPRGNIOGKTLVSQLELQDSG 180  
 DB 121 LVFGLTANSDPHLLQGSGLTLTLESPGSSPSVQCRSPRGNIOGKTLVSQLELQDSG 180  
 QY 181 TWCTVLQGNKKVEFKIDIVLAFAQASSIYKKEGEVEFSFPLAFTVEKLTGSGELMW 240  
 DB 181 TWCTVLQGNKKVEFKIDIVLAFAQASSIYKKEGEVEFSFPLAFTVEKLTGSGELMW 240  
 QY 241 QAEPRSSSKSWITPPLKKEVSVKRVTDPKLQMGKPLPHLTLPOALPOYAGSGNLTLA 300  
 DB 241 QAEPRSSSKSWITPPLKKEVSVKRVTDPKLQMGKPLPHLTLPOALPOYAGSGNLTLA 300  
 QY 301 LEAKGKGLHGVNVLVWRATOLQKVLTCCEVWGPTSPKMLSLKLENKEAKVSKREKPYWV 360  
 DB 301 LEAKGKGLHGVNVLVWRATOLQKVLTCCEVWGPTSPKMLSLKLENKEAKVSKREKPYWV 360  
 QY 361 LNPEAGMOCCLSDSGVLLSNIKVLPTWSTPVHPRASALPAPPTGSALPDPOYASALP 420  
 DB 361 LNPEAGMOCCLSDSGVLLSNIKVLPTWSTPVHPRASALPAPPTGSALPDPOYASALP 420  
 QY 421 DPPAASALPALAVISFLIGLGLV-ACVLAATR 453  
 DB 421 DPPAASALPALAVISFLIGLGLV-ACVLAATR 453  
 QY 397 -----MGLIVLGVAGLLFLTGLGIFCVCRHR 425  
 DB 397 -----MGLIVLGVAGLLFLTGLGIFCVCRHR 425

RESULT 71

AAAP90833 standard; protein; 384 AA.

AC AAP90833;

DT 25-MAR-2003 (revised)

DT 01-AUG-1990 (first entry)

DE Amino acid sequence of a soluble T4-like (srT4) polypeptide encoded by a

DE portion of clone p159-7 (PL mutet. srT4).

KW HIV, immunotherapeutic; prophylactic; soluble T4-like polypeptide;

KW diagnostic; p159-7 (PL mutet. srT4).

OS Homo sapiens.

FT Key Location/Qualifiers

FT Misc-difference 10 /note= "Initiating Met"

PN W08901940-A.

PD 09-MAR-1989.

PF 01-SEP-1988; 88WO-US002940.

PR 04-SEP-1987; 87US-00094322.  
 PR 07-JAN-1988; 88US-00141649.  
 XX (BIOJ) BIOGEN INC.  
 PA (BIOJ) BIOGEN INC.  
 XX  
 PI Fisher RA, Gilbert W, Sato VL, Flavell RA, Maraganore JM;  
 XX  
 XX WPI; 1989-085519/11.  
 DR N-PSDB; AAN90643.  
 XX  
 PT DNA sequences coding for soluble T4-like polypeptide(s) - used in  
 PT immuno-therapeutic and immunosuppressive comps. and for preventing,  
 PT treating or detecting AIDS.  
 XX  
 XX Disclosure; Page 7; 207pp; English.  
 CC It is the protein sequence encoded by the srT4 sequence. It is claimed in  
 CC the patent. It is useful in immunotherapeutic, prophylactic and  
 CC diagnostic comps. It can be used to purify HIV from a sample. (Updated  
 CC on 25-MAR-2003 to correct PR field.)  
 XX  
 SO Sequence 384 AA;  
 Query Match 82.4%; Score 1915; DB 1; Length 384;  
 Best Local Similarity 99.3%; Pred. No. 1.2e-122;  
 Matches 371; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 24 QGNKVVLGKGGDTVELCTASQKSIQFHMKNNSQIKILGNQSFLLTKGPKSLNDRADSR 83  
 DB 11 QGNKVVLGKGGDTVELCTASQKSIQFHMKNNSQIKILGNQSFLLTKGPKSLNDRADSR 70  
 QY 84 RSLMDQGNFPLIIKMLKIEDSDTYICEVEDQKEEYQLLVGLTANSDPHLLQGSGLTLTL 143  
 DB 71 RSLMDQGNFPLIIKMLKIEDSDTYICEVEDQKEEYQLLVGLTANSDPHLLQGSGLTLTL 130  
 QY 144 ESPPGSSPSVQCRSPRGNIOGKTLVSQLELQDSGTWCTVLQGNKKVEFKIDIVLA 203  
 DB 131 ESPPGSSPSVQCRSPRGNIOGKTLVSQLELQDSGTWCTVLQGNKKVEFKIDIVLA 190  
 QY 204 FORASSIYKKEGEVEFSFPLAFTVEKLTGSGELMWQAEPRSSSKSWITFDLKNKEVSV 263  
 DB 191 FORASSIYKKEGEVEFSFPLAFTVEKLTGSGELMWQAEPRSSSKSWITFDLKNKEVSV 250  
 QY 264 KRYTODPKLQMGKPLPHLTLPOALPOYAGSGNLTALBAKTGLHGVNVLVWRATOLQ 323  
 DB 251 KRYTODPKLQMGKPLPHLTLPOALPOYAGSGNLTALBAKTGLHGVNVLVWRATOLQ 310  
 QY 324 KNLTCCEVWGPTSPKMLSLKLENKEAKVSKREKPYWVLNPEAGMOCCLSDSGVLLSNI 383  
 DB 311 KNLTCCEVWGPTSPKMLSLKLENKEAKVSKREKPYWVLNPEAGMOCCLSDSGVLLSNI 370  
 QY 384 IKVLPWSTPVHP 396  
 DB 371 IKVLPWSTPVHP 383  
 RESULT 72  
 AAR41041  
 ID AAR41041 standard; protein; 942 AA.  
 AC AAR41041;  
 XX  
 XX  
 DT 24-OCT-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 22-MAR-1994 (first entry)  
 XX  
 DE CD4-GBP130 fusion protein.  
 XX  
 KW Merozoite; Glycophorin Binding Protein 130; malaria; HIV; env;  
 KW human immunodeficiency virus; envelope glycoprotein; hybrid protein;  
 KW red blood cell; erythrocyte; AIDS.  
 XX  
 OS Homo sapiens.

```

OS Plasmidium falciparum.
OS Chimeric.
FH Key Location/Qualifiers
FT Region 1..371
FT Region /note="Residues 1-371 of CD4"
FT Region 372..942
FT Region /note="Residues 201-774 of GBP130"
XX
XX MO9318160-A1.
XX
XX 16-SEP-1993.
XX
XX 10-MAR-1993; 93WO-GB000505.
XX
XX 11-MAR-1992; 92GB-00005276.
XX 08-JUL-1992; 92GB-00014481.
XX 24-JUL-1992; 92GB-00015829.
XX 16-SEP-1992; 92GB-00019562.
XX 03-MAR-1993; 93GB-00004311.
XX
XX (PREN/) PRENDERGAST K F.
XX
XX Prendergast KF;
XX
XX WPI; 1993-303474/38.
XX
XX Anti-viral fusion peptide(s) - comprise viral-binding component and
XX malaria merozoite red cell binding component, for treating e.g. HIV, and
XX hepatitis.
XX
XX Claim 2; Page 35-37; 69pp; English.
XX
XX The hybrid protein NH2-CD4(1-371)-GBP130(201-774)-COOH is a specifically
XX claimed example of a fusion protein; it comprises at
XX least part of the CD4 molecule fused to a peptide from a malarial
XX parasite merozoite protein with affinity for red blood cells. The fusion
XX protein can bind free HIV in the blood to red blood cells and
XX consequently reduce viral titre, prevent transmission of the virus and
XX improve safety of blood transfusions. (Updated on 25-MAR-2003 to correct
XX PN field.) (Updated on 24-OCT-2003 to standardise OS field)
XX
XX Sequence 942 AA;
XX
Query Match 82.2%; Score 1911; DB 2; Length 942;
Best Local Similarity 98.7%; Pred. No. 6,9e-122;
Matches 371; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 24 QGNKVVLAGKKDPTVELTCTASQKKSIOFHMKNSNOIKILGQSGFLTGPSPKLNDRADR 83
DB 1 QGNKVVLAGKKDPTVELTCTASQKKSIOFHMKNSNOIKILGQSGFLTGPSPKLNDRADR 60
QY 84 RSLMDGNGPPLIINKLIKEDSDTYICEVEDOKEEVQLVFGTLTANSDTLHLOGSILTLTL 143
DB 61 RSLMDGNGPPLIINKLIKEDSDTYICEVEDOKEEVQLVFGTLTANSDTLHLOGSILTLTL 120
QY 144 EEPGSSPSVQCRSPRGKNIQSGKITSVSQLELDSDGTCTVLONQKVEFKIDIVLA 203
DB 121 EEPGSSPSVQCRSPRGKNIQSGKITSVSQLELDSDGTCTVLONQKVEFKIDIVLA 180
QY 204 FOKASSIVKKEGEOVEFSFPLAFTVEKLTGSGELMWMQERASSSSSWITPFLKNEVEV 263
DB 181 FOKASSIVKKEGEOVEFSFPLAFTVEKLTGSGELMWMQERASSSSSWITPFLKNEVEV 240
QY 264 KEVTDPKQLQMGKULPLHLTLPOALPOVAGSGNLTALBAKTGKLEHVNLVMMRATOLQ 323
DB 241 KEVTDPKQLQMGKULPLHLTLPOALPOVAGSGNLTALBAKTGKLEHVNLVMMRATOLQ 300
QY 324 KULTCEVMGPTSPKMLSLKLENKAKVSKREKPVWLVNPEAGMOCCLSDSGOVLBSN 383
DB 301 KULTCEVMGPTSPKMLSLKLENKAKVSKREKPVWLVNPEAGMOCCLSDSGOVLBSN 360
QY 384 IKVLPTWSTPVPVHPRAS 399

```

```

DB 361 IKVLPTWSTPVPVHPRAS 376
|||||
RESULT 73
AAR41043
ID AAR41043 standard; protein; 1786 AA.
XX
XX AAR41043;
XX
XX 24-OCT-2003 (revised)
XX 25-MAR-2003 (revised)
XX 22-MAR-1994 (first entry)
XX
XX CD4-EBA175 fusion protein.
XX
XX Merozoite; Erythrocyte Binding Antigen 175; malaria; HIV; env;
XX human immunodeficiency virus; envelope glycoprotein; hybrid protein;
XX red blood cell; erythrocyte; AIDS; molecular machine.
XX
XX Homo sapiens.
XX Plasmidium falciparum.
XX Chimeric.
FH Key Location/Qualifiers
FT Region 1..371
FT Region /note="Residues 1-371 of CD4"
FT Region 372..1786
FT Region /note="Residues 20-1435 of EBA-175"
XX
XX MO9318160-A1.
XX
XX 16-SEP-1993.
XX
XX 10-MAR-1993; 93WO-GB000505.
XX
XX 11-MAR-1992; 92GB-00005276.
XX 08-JUL-1992; 92GB-00014481.
XX 24-JUL-1992; 92GB-00015829.
XX 16-SEP-1992; 92GB-00019562.
XX 03-MAR-1993; 93GB-00004311.
XX
XX (PREN/) PRENDERGAST K F.
XX
XX Prendergast KF;
XX
XX WPI; 1993-303474/38.
XX
XX Anti-viral fusion peptide(s) - comprise viral-binding component and
XX malaria merozoite red cell binding component, for treating e.g. HIV, and
XX hepatitis.
XX
XX Claim 9; Page 44-47; 69pp; English.
XX
XX The hybrid protein NH2-CD4(1-371)-EBA175(20-1435)-COOH is a specifically
XX claimed example of a fusion protein; it comprises at
XX least part of the CD4 molecule fused to a peptide from a malarial
XX parasite merozoite protein with affinity for red blood cells. The fusion
XX protein can bind free HIV in the blood to red blood cells and
XX consequently reduce viral titre, prevent transmission of the virus and
XX improve safety of blood transfusions. (Updated on 25-MAR-2003 to correct
XX PN field.) (Updated on 24-OCT-2003 to standardise OS field)
XX
XX Sequence 1786 AA;
XX
Query Match 82.2%; Score 1911; DB 2; Length 1786;
Best Local Similarity 99.2%; Pred. No. 1.5e-121;
Matches 371; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 24 QGNKVVLAGKKDPTVELTCTASQKKSIOFHMKNSNOIKILGQSGFLTGPSPKLNDRADR 83
DB 1 QGNKVVLAGKKDPTVELTCTASQKKSIOFHMKNSNOIKILGQSGFLTGPSPKLNDRADR 60

```

QY 84 RSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOLLVFGLTANSDFHLLQGOSLTTLT 143  
 DB 61 RSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOLLVFGLTANSDFHLLQGOSLTTLT 120  
 QY 144 ESPSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONQKVEFKIDIVLA 203  
 DB 121 ESPSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONQKVEFKIDIVLA 180  
 QY 204 FOKASSIYKKEGEOVESFPPLAFVTEKLTSGELMWOAERASSSSKSWITFDLNKEVSV 263  
 DB 181 FOKASSIYKKEGEOVESFPPLAFVTEKLTSGELMWOAERASSSSKSWITFDLNKEVSV 240  
 QY 264 KRVTDPRKQMGKKPLHLTLPOALPOYAGSNTLTALEAKTGKHQEVNIVVMRATOLQ 323  
 DB 241 KRVTDPRKQMGKKPLHLTLPOALPOYAGSNTLTALEAKTGKHQEVNIVVMRATOLQ 300  
 QY 324 KNLTEVWGPTSPKMLSLKENKEAKVSKREKPVWVLPNPAQMOCCLSDSGOVLLESN 383  
 DB 301 KNLTEVWGPTSPKMLSLKENKEAKVSKREKPVWVLPNPAQMOCCLSDSGOVLLESN 360  
 QY 384 IKVLPTWSTPVNPR 397  
 DB 361 IKVLPTWSTPVNPR 374

RESULT 74  
 AAM41376  
 ID AAM41376 standard; peptide; 433 AA.

AC AAM41376;  
 DT 28-MAY-1998 (first entry)  
 DE Human CD4.  
 XX Antibody; CD4; passive immunity; HIV type 1; HIV type 2; HIV infection;  
 KM simian immunodeficiency virus; SIV; AIDS; therapy; HIV gp120.  
 XX Homo sapiens.  
 PN MO9746697-A2.  
 PD 11-DEC-1997.  
 XX 03-JUN-1997; 97WO-US009449.  
 PF 03-JUN-1996; 96US-00657149.  
 PR 28-FEB-1997; 97US-00808374.  
 PR 02-JUN-1997; 97US-00867149.  
 XX (UNBI-) UNITED BIOMEDICAL INC.  
 PA Wang CY;  
 PI WPI. 1998-042204/04.  
 DR WPI. 1998-042204/04.  
 XX Antibody against complex of CD4 and chemokine receptor domain - useful  
 PT for prevention and treatment of human immunodeficiency virus infection.  
 XX Disclosure: Page 122-123; 140pp; English.

CC This sequence represents human CD4, and was used to isolate the antibody  
 CC (Ab) of the invention. The Ab, preferably a M2 or B13 Ab, has the  
 CC following characteristics: (a) binds recp4; (b) binds CD4 expressing  
 CC cells in an immunofluorescence assay, where the binding pattern is in the  
 CC shape of "caps" when examined with a high resolution fluorescence  
 CC microscope; (c) blocks the binding of HIV glycoprotein 120 (gp120) to CD4  
 CC expressing cells; (d) binds CD4 expressing cells previously bound with  
 CC gp120; and (e) neutralises HIV primary isolates in an in vitro  
 CC microplaque assay at a concentration of less than 10 mu g/ml, preferably  
 CC at a concentration in the range of 0.01-10 mu g/ml for 50% neutralisation  
 CC and 0.1-35 mu g/ml for 90% neutralisation. The Ab can be used to provide  
 CC passive immunity to HIV in a mammal, when administered parenterally,

CC specifically all clades of HIV type 1, and from diverse primary isolates  
 CC of HIV type 2 and simian immunodeficiency virus (SIV). The Ab is  
 CC prophylactic and therapeutic for HIV infection and all stages of AIDS  
 CC because it prevents replicative infection of host cells both before and  
 CC after HIV gp120 has bound to the host cell antigen complex comprising CD4  
 CC on the surface of CD4 positive lymphocytes, thus it is capable of  
 CC preventing HIV infection and retarding the spread of the virus to  
 CC uninfected cells. It is also uniquely useful because it inhibits  
 CC infection following binding of HIV to CD4 expressing cells

XX Sequence 433 AA;

QY Query Match 82.1%; Score 1909; DB 2; Length 433;  
 DB Best Local Similarity 88.6%; Pred. No. 3.6e-122;  
 DB Matches 380; Conservative 2; Mismatches 17; Indels 30; Gaps 2;

QY 26 NKVVLGKGGDTVELTCTASQKSIQFHWKMNQIKILGNGSFLTKPSKLNDRADRRS 85  
 DB 1 NKVVLGKGGDTVELTCTASQKSIQFHWKMNQIKILGNGSFLTKPSKLNDRADRRS 60  
 QY 86 LMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOLLVFGLTANSDFHLLQGOSLTTLTLES 145  
 DB 61 LMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOLLVFGLTANSDFHLLQGOSLTTLTLES 120  
 QY 146 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONQKVEFKIDIVLAFO 205  
 DB 121 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONQKVEFKIDIVLAFO 180  
 QY 206 KASSIYKKEGEOVESFPPLAFVTEKLTSGELMWOAERASSSSKSWITFDLNKEVSVKR 265  
 DB 181 KASSIYKKEGEOVESFPPLAFVTEKLTSGELMWOAERASSSSKSWITFDLNKEVSVKR 240  
 QY 266 VTQDPKQMGKKPLHLTLPOALPOYAGSNTLTALEAKTGKHQEVNIVVMRATOLQKN 325  
 DB 241 VTQDPKQMGKKPLHLTLPOALPOYAGSNTLTALEAKTGKHQEVNIVVMRATOLQKN 300  
 QY 326 LTCFVWGPTSPKMLSLKENKEAKVSKREKPVWVLPNPAQMOCCLSDSGOVLLESNIK 385  
 DB 301 LTCFVWGPTSPKMLSLKENKEAKVSKREKPVWVLPNPAQMOCCLSDSGOVLLESNIK 360  
 QY 386 VLPTWSTPVNPRASALPAPPTGSALPDQFASALPDPPASALPALAVISFLLGLGLGV 445  
 DB 361 VLPTWSTPVNPRASALPAPPTGSALPDQFASALPDPPASALPALAVISFLLGLGLGI 391  
 QY 446 -ACVLTART 453  
 DB 392 PFCVRCRHR 400

RESULT 75  
 AAR41042  
 ID AAR41042 standard; protein; 729 AA.

XX AAR41042;  
 AC AAR41042;  
 XX 24-OCT-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 22-MAR-1994 (first entry)  
 XX CD4-GBPH fusion protein.

XX Merozoite; Glycophorin Binding Protein homologue; malaria; HIV; env;  
 KW human immunodeficiency virus; envelope glycoprotein; hybrid protein;  
 KW red blood cell; erythrocyte; AIDS.

OS Homo sapiens.  
 OS Plasmodium falciparum.  
 OS Chimeric.

XX Key Location/Qualifiers  
 FT Region 1..371  
 FT /note="residues 1-371 of CD4"  
 FT Region 372..729



FT /note= "residues 70-427 of GBPH"  
XX  
XX WO318160-A1.  
XX  
XX 16-SEP-1993.  
XX  
XX 10-MAR-1993; 93WO-GB000505.  
XX  
XX 11-MAR-1992; 92GB-00005276.  
XX PR 08-JUL-1992; 92GB-00014481.  
XX PR 24-JUL-1992; 92GB-00015829.  
XX PR 16-SEP-1992; 92GB-00019562.  
XX PR 03-MAR-1993; 93GB-00004311.  
XX  
XX (PREN/) PRENDERGAST K F.  
XX  
XX Prendergast KF;  
XX  
XX WPI; 1993-303474/38.  
XX  
XX Anti-viral fusion peptide(s) - comprise viral-binding component and  
XX PT malaria merozoite red cell binding component, for treating e.g. HIV, and  
XX PT hepatitis.  
XX  
XX Claim 7; Page 40-41; 69pp; English.  
XX  
XX The hybrid protein NM2-CD4(1-371)-GBPH(70-427)-COOH is a specifically  
XX CC claimed example of a fusion protein of the invention; it comprises at  
XX CC least part of the CD4 molecule fused to a peptide from a malarial  
XX CC parasite merozoite protein with affinity for red blood cells. The fusion  
XX CC protein can bind free HIV in the blood to red blood cells and  
XX CC consequently reduce viral titre, prevent transmission of the virus and  
XX CC improve safety of blood transfusions. (Updated on 25-MAR-2003 to correct  
XX CC PN field.) (Updated on 24-OCT-2003 to standardise OS field)  
SQ Sequence 729 AA;  
  
Query Match 82.1%; Score 1908; DB 2; Length 729;  
Best Local Similarity 99.7%; Pred. No. 8e-122;  
Matches 370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 24 QGNKVVLGKKGDVVELTCTASQKKSIOFHMKNNSQIKILGNQGSFLTGGPSKLNDRADSR 83  
DB 1 QGNKVVLGKKGDVVELTCTASQKKSIOFHMKNNSQIKILGNQGSFLTGGPSKLNDRADSR 60  
  
QY 84 RSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLQGSILTLTL 143  
DB 61 RSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLQGSILTLTL 120  
  
QY 144 EEPGSSPSVQCRSPRGKNIQGGKTLSSVQLLEQDSGTWCTVLOQKVEFKIDIVVLAFO 205  
DB 121 EEPGSSPSVQCRSPRGKNIQGGKTLSSVQLLEQDSGTWCTVLOQKVEFKIDIVVLAFO 180  
  
QY 204 POKASSIYVKKEGQVEFFSFPPLAFYVEKLTGSGELMMQAEARASSSSSWITFDLKNKEVSYKR 265  
DB 181 POKASSIYVKKEGQVEFFSFPPLAFYVEKLTGSGELMMQAEARASSSSSWITFDLKNKEVSYKR 240  
  
QY 264 KRVTDOPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKLHQEVNLYVMRATOLQKN 325  
DB 241 KRVTDOPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKLHQEVNLYVMRATOLQKN 300  
  
QY 324 KULTCVWGPSTSPKMLSLKLENKEAKVSKREKPVWVYLNPEAGMOCCLSDSGOVLLESNIK 385  
DB 301 KULTCVWGPSTSPKMLSLKLENKEAKVSKREKPVWVYLNPEAGMOCCLSDSGOVLLESNIK 360  
  
QY 384 IKVLPWTSTPVP 394  
DB 361 IKVLPWTSTPVP 371

XX  
XX AAR07721;  
XX AC  
XX 25-MAR-2003 (revised)  
XX DT 18-FEB-1991 (first entry)  
XX DT  
XX DE Recombinant soluble (rs) T4.  
XX XX  
XX KM Metal-binding site.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO9012803-A.  
XX  
XX PD 01-NOV-1990.  
XX  
XX PF 14-APR-1989; 89US-00338991.  
XX  
XX PR 14-APR-1989; 89US-00338991.  
XX  
XX PA (BIOJ ) BIOGEN INC.  
XX PA (BIOJ ) BIOGEN INC.  
XX  
XX PI Staples MA, Pargellis CA;  
XX  
XX DR WPI; 1990-348421/46.  
XX  
XX PT Purifying protein having surface metal-binding amino acid residues -  
XX PT using an immobilised metal affinity chromatography resin.  
XX  
XX PS Disclosure; Fig 2; 36pp; E.  
XX  
XX CC The rs T4 can be purified from a crude sample (Updated on 25-MAR-2003 to  
XX CC correct PA field.)  
XX  
SQ Sequence 375 AA;  
  
Query Match 81.9%; Score 1904; DB 2; Length 375;  
Best Local Similarity 98.9%; Pred. No. 6.7e-122;  
Matches 369; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
  
QY 26 NKVVVGKKGDVVELTCTASQKKSIOFHMKNNSQIKILGNQGSFLTGGPSKLNDRADSRRS 85  
DB 1 NKVVVGKKGDVVELTCTASQKKSIOFHMKNNSQIKILGNQGSFLTGGPSKLNDRADSRRS 60  
  
QY 86 LMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLQGSILTLTL 145  
DB 61 LMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLQGSILTLTL 120  
  
QY 146 PPGSSPSVQCRSPRGKNIQGGKTLSSVQLLEQDSGTWCTVLOQKVEFKIDIVVLAFO 205  
DB 121 PPGSSPSVQCRSPRGKNIQGGKTLSSVQLLEQDSGTWCTVLOQKVEFKIDIVVLAFO 180  
  
QY 206 KASSIYVKKEGQVEFFSFPPLAFYVEKLTGSGELMMQAEARASSSSSWITFDLKNKEVSYKR 265  
DB 181 KASSIYVKKEGQVEFFSFPPLAFYVEKLTGSGELMMQAEARASSSSSWITFDLKNKEVSYKR 240  
  
QY 266 VTQDBKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKLHQEVNLYVMRATOLQKN 325  
DB 241 VTQDBKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKLHQEVNLYVMRATOLQKN 300  
  
QY 326 LTCVWGPSTSPKMLSLKLENKEAKVSKREKPVWVYLNPEAGMOCCLSDSGOVLLESNIK 385  
DB 301 LTCVWGPSTSPKMLSLKLENKEAKVSKREKPVWVYLNPEAGMOCCLSDSGOVLLESNIK 360  
  
QY 386 VLPTWSTPVP 398  
DB 361 VLPTWSTPVP 373

RESULT 76  
AAR07721  
ID AAR07721 standard; protein; 375 AA.

RESULT 77  
AA54500  
ID AA54500 standard; protein; 433 AA.

```

XX AC AAY54500;
XX DT 25-APR-2000 (first entry)
XX DE Amino acid sequence of the human CD4 protein.
XX
XX Human; CD4 protein; antigenic peptide; CDR2-like domain; apoptosis;
XX syncytia formation; human immune deficiency virus; HIV binding;
XX CD4-Class II interaction; immunisation; CD4 surface complex;
XX immune response; transplant rejection; autoimmune diseases;
XX rheumatoid arthritis; systemic lupus erythematosus; psoriasis.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Domain 27..66
XX /note="CDR-2 like domain; specifically claimed in claim
XX 1"
XX
XX WO967294-A1.
XX
XX 29-DEC-1999.
XX
XX 21-JUN-1999; 99WO-US014030.
XX
XX 20-JUN-1998; 98US-00100409.
XX
XX (UNBI-) UNITED BIOMEDICAL INC.
XX
XX Wang CY;
XX
XX WPI; 2000-160579/14.
XX
XX New antigenic peptide from the CDR2 domain of CD4, for immunization
XX against e.g. human immune deficiency virus.
XX
XX Claim 1; Page 70-71; 106pp; English.
XX
XX The present sequence represents the human CD4 protein. The specification
XX describes antigenic peptides derived from the CDR2-like domain of CD4
XX (amino acids 27-66 of AAY54500). These antigenic peptides present
XX neutralising receptor/co-receptor effector sites of the CDR2-like domain.
XX The peptides evoke effective antibody responses by having optimised site-
XX specificity. The induced antibodies block human immune deficiency virus
XX (HIV) binding and syncytia formation. They may also block CD4-Class II
XX interactions with other cells, deliver signals to T cells (inhibiting
XX normal CD4-mediated immunoregulatory functions) or induce apoptosis of
XX CD4 cells by simultaneous engagement of T cell receptors. Conjugates and
XX peptides containing the antigenic peptides are used for active
XX immunisation to generate antibodies against CD4 surface complexes,
XX especially to prevent binding of HIV to CD4 and thus HIV infection, but
XX also to treat undesirable immune responses such as transplant rejection,
XX or autoimmune diseases (rheumatoid arthritis, systemic lupus
XX erythematosus or psoriasis). These conjugates produce high-titre
XX antibodies which are broadly neutralising against primary isolates from
XX all classes of HIV-1 and of HIV-2. The peptides may be cyclically
XX constructed and may include a promiscuous T helper epitope that is active
XX in genetically diverse subjects
XX
XX Sequence 433 AA;
XX
XX Query Match 81.9%; Score 1904; DB 3; Length 433;
XX Best Local Similarity 88.6%; Pred. No. 8e-122;
XX Matches 379; Conservative 2; Mismatches 17; Indels 30; Gaps 2;
XX
XX 27 KVLGKQDYEELTCTASQKSIQPHWKNQIKILNGQSFLLTKGPSKLNDRADSRSL 86
XX 2 KVLGKQDYEELTCTASQKSIQPHWKNQIKILNGQSFLLTKGPSKLNDRADSRSL 61
XX 87 WDQGNFLLIKLKIEDSDTYICEVEDQKEVQLLVFGLTANSDTHLLQGGSLTTLTLESP 146
XX 62 WDQGNFLLIKLKIEDSDTYICEVEDQKEVQLLVFGLTANSDTHLLQGGSLTTLTLESP 121

```

```

QY 147 PGSSPSVQCRSPRGKNIQGGKTLISVSQLELDSDGTWCTVLQNOQKVEFKIDIVLAFQK 206
DB 122 PGSSPSVQCRSPRGKNIQGGKTLISVSQLELDSDGTWCTVLQNOQKVEFKIDIVLAFQK 181
QY 207 ASSIVYKKEGEQVFEFPLPAFTVEKLTGSGELMWQERASSSKSWITFDLKNKVSYKRV 266
DB 182 ASSIVYKKEGEQVFEFPLPAFTVEKLTGSGELMWQERASSSKSWITFDLKNKVSYKRV 241
QY 267 TODPKLQMGKPLPLHLTLPPALPOYAGSGLTLALFAKTKLHOEVNLVVMRATQQLKNL 326
DB 242 TODPKLQMGKPLPLHLTLPPALPOYAGSGLTLALFAKTKLHOEVNLVVMRATQQLKNL 301
QY 327 TCEWVGPTSPKMLSLKENKEAKVSRKRPVWVLNPEAGMWQCLSDSGVLESNIKV 386
DB 302 TCEWVGPTSPKMLSLKENKEAKVSRKRPVWVLNPEAGMWQCLSDSGVLESNIKV 361
QY 387 LPTWSTPVHPRASALPAPPTGSALPDPTNASALPDPPASALPAAALAVISFLGLGLGV- 445
DB 362 LPTWSTPVQP-----MALIVLGAGVALLFTIGLIG 392
QY 446 ACVLARTR 453
DB 393 FCVRCRRR 400
XX
XX RESULT 78
XX AAP6151
XX ID AAP6151 standard; protein; 434 AA.
XX AC AAP6151;
XX AD 25-MAR-2003 (revised)
XX DT 24-DEC-1990 (first entry)
XX DE Sequence of a fusion of the herpes virus and N-terminal 27 residues to
XX DE the putative mature N-terminus of CD4T.
XX KM gp120; HIV; therapy; adhesion; antiviral.
XX XX Synthetic.
XX OS EP14317-A.
XX FN 03-MAY-1989.
XX PD 03-OCT-1988; 88EP-00309194.
XX PE 02-OCT-1987; 87US-00104329.
XX PR 28-SEP-1988; 88US-00250785.
XX XX (GERTH) GENENTECH INC.
XX PA Capon DJ, Gregory TU;
XX PI WPI; 1989-131855/18.
XX DR N-PSDB; AAN90778.
XX XX Compars. contg. adhesion variants - useful in therapy and diagnostics,
XX PT e.g. CD4 variants which are therapeutically useful for treating human
XX PT immune-deficiency virus.
XX XX Example; Fig 2a-2c; 36pp; English.
XX XX A nucleic acid encoding an AA sequence variant of an adhesion is claimed.
XX CC Adhesion variants may be a fusion of a CD4 polypeptide and a polypeptide
XX CC different from CD4. Used therapeutically or as diagnostic reagents for
XX CC the assay of adhesions or their ligands, or for purificn. of the adhesions.
XX CC The CD4 adhesion variants are used for antiviral or immunomodulatory
XX CC therapy, in partic. treatment of HIV infection. (Updated on 25-MAR-2003
XX CC to correct FI field.)
XX
XX Sequence 434 AA;
XX

```

Query Match 81.9%; Score 1904; DB 1; Length 434;  
Best Local Similarity 99.7%; Pred. No. 8e-122;  
Matches 369; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 24 QGNKVVLGKKGDTVELTCTASQKSIQPHWKNSNOIKILGNQGSFLTGKPSKLNDRADSR 83  
DB 56 QGNKVVLGKKGDTVELTCTASQKSIQPHWKNSNOIKILGNQGSFLTGKPSKLNDRADSR 115

QY 84 RSLMDQGNPPLIIKMLKIEDSDTYICEVEDQKEEVQLVFGILTANSDTHLQOGSITLT 143  
DB 116 RSLMDQGNPPLIIKMLKIEDSDTYICEVEDQKEEVQLVFGILTANSDTHLQOGSITLT 175

QY 144 ESPSPSSSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNOKKVEFKIDIVVLA 203  
DB 176 ESPSPSSSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNOKKVEFKIDIVVLA 235

QY 204 FOKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMWAERASSSKSWITFDLKNKEVS 263  
DB 236 FOKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMWAERASSSKSWITFDLKNKEVS 295

QY 264 KRYTODPKLQMGKPLHLTLPOALPOYAGSGNLTALFAKTGKLGHOEVNLVYMRATQ 323  
DB 296 KRYTODPKLQMGKPLHLTLPOALPOYAGSGNLTALFAKTGKLGHOEVNLVYMRATQ 355

QY 324 KNLTCVWGPSTSPKMLSLKLENKEAKVSKREKPVVNLNPEAGMOCCLSDSGQVLTESN 383  
DB 356 KNLTCVWGPSTSPKMLSLKLENKEAKVSKREKPVVNLNPEAGMOCCLSDSGQVLTESN 415

QY 384 IKVLPTWSTP 393  
DB 416 IKVLPTWSTP 425

RESULT 79  
AAP93557  
ID AAP93557 standard; protein; 434 AA.  
XX  
AC AAP93557;  
XX  
XX 25-MAR-2003 (revised)  
DT 03-OCT-2002 (revised)  
DT 06-JUN-1990 (first entry)  
XX  
DE Fusion of the herpes gD leader and N-terminal 27 residues to mature N-terminalus of CD4T.  
XX  
XX Adhesion variant; CD4T; soluble CD4; truncated CD4; antiviral; HIV;  
KM herpes gD; immunomodulatory; diagnostic.  
XX  
OS Homo sapiens.  
XX  
XX WO8902922-A.  
PN  
XX  
XX 06-APR-1989.  
PD  
XX  
XX 03-OCT-1988; 88WO-US003414.  
PF  
XX  
XX 02-OCT-1987; 87US-00104329.  
PR 28-SEP-1988; 88US-00250785.  
XX  
XX (GETH ) GENENTECH INC.  
PA  
XX  
XX Capon DJ, Gregory TJ;  
FI  
XX  
XX WPI; 1989-114397/15.  
DR  
XX  
XX N-PSDB; AAN90735.  
XX  
XX New nucleic acid sequences encoding adhesion, esp. CD 4, variants -  
PT partic. with trans-membrane domain inactivated or fused to other peptide,  
PT useful esp. for treating HIV infections.  
XX  
XX  
XX  
PS Fig 2A-2C; pp. 5/13-8/13; 78pp; English.

XX  
CC CD4T is a truncated or soluble variant of CD4. CD4T fusion proteins can  
CC have antiviral and immunomodulatory activity and are esp. useful for  
CC treating HIV infections, regardless of genetic variation within the  
CC virus. CD4T fusion proteins, and antibodies raised against them, can also  
CC be used diagnostically for assaying adhesion and their ligands. (Updated  
CC on 03-OCT-2002 to add missing OS field.) (Updated on 25-MAR-2003 to  
CC correct PR field.) (Updated on 25-MAR-2003 to correct PA field.)  
XX  
SQ Sequence 434 AA;  
Query Match 81.9%; Score 1904; DB 1; Length 434;  
Best Local Similarity 99.7%; Pred. No. 8e-122;  
Matches 369; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 24 QGNKVVLGKKGDTVELTCTASQKSIQPHWKNSNOIKILGNQGSFLTGKPSKLNDRADSR 83  
DB 56 QGNKVVLGKKGDTVELTCTASQKSIQPHWKNSNOIKILGNQGSFLTGKPSKLNDRADSR 115

QY 84 RSLMDQGNPPLIIKMLKIEDSDTYICEVEDQKEEVQLVFGILTANSDTHLQOGSITLT 143  
DB 116 RSLMDQGNPPLIIKMLKIEDSDTYICEVEDQKEEVQLVFGILTANSDTHLQOGSITLT 175

QY 144 ESPSPSSSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNOKKVEFKIDIVVLA 203  
DB 176 ESPSPSSSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNOKKVEFKIDIVVLA 235

QY 204 FOKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMWAERASSSKSWITFDLKNKEVS 263  
DB 236 FOKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMWAERASSSKSWITFDLKNKEVS 295

QY 264 KRYTODPKLQMGKPLHLTLPOALPOYAGSGNLTALFAKTGKLGHOEVNLVYMRATQ 323  
DB 296 KRYTODPKLQMGKPLHLTLPOALPOYAGSGNLTALFAKTGKLGHOEVNLVYMRATQ 355

QY 324 KNLTCVWGPSTSPKMLSLKLENKEAKVSKREKPVVNLNPEAGMOCCLSDSGQVLTESN 383  
DB 356 KNLTCVWGPSTSPKMLSLKLENKEAKVSKREKPVVNLNPEAGMOCCLSDSGQVLTESN 415

QY 384 IKVLPTWSTP 393  
DB 416 IKVLPTWSTP 425

RESULT 80  
AAR74222  
ID AAR74222 standard; protein; 432 AA.  
XX  
AC AAR74222;  
XX  
XX 25-MAR-2003 (revised)  
DT 26-NOV-1995 (first entry)  
XX  
DE Epitope on the primary CD4 sequence.  
XX  
XX Chimaeitic; mutant; mapping; immunodiagnostics.  
KM  
XX  
XX Synthetic.  
OS  
XX  
XX US5411861-A.  
PN  
XX  
XX 02-MAY-1995.  
PD  
XX  
XX 27-FEB-1992; 92US-00842465.  
PF  
XX  
XX 15-APR-1988; 88US-00181826.  
PR  
XX  
XX (GEMO ) GEN HOSPITAL CORP.  
PA  
XX  
XX Seed B, Peterson A;  
PI  
XX  
XX  
XX  
XX WPI; 1995-178122/23.

PT Mutational analysis method for protein epitope(s) - by expressing mutant  
 PT cDNA and using negative and positive selection to identify binding loss  
 PT mutants.

PS Disclosure: Fig 2, 28pp; English.

CC The sequence is that of an epitope on the primary CD4 sequence. The  
 CC epitope sequence can be identified by vector pIH3M. The See also  
 CC AAR74221. (Updated on 25-MAR-2003 to correct PF field.)

XX Sequence 432 AA;

Query Match 81.8%; Score 1901; DB 2; Length 432;

Best Local Similarity 88.6%; Pred. No. 1.3e-121;  
 Matches 379; Conservative 2; Mismatches 17; Indels 30; Gaps 2;

QY 27 KVLVGGKDPVLTCTTASQKSIQPHMKNSNOIKILGNQGSFLTKGPKLNDRADSRSL 86  
 DB 1 KVLVGGKDPVLTCTTASQKSIQPHMKNSNOIKILGNQGSFLTKGPKLNDRADSRSL 60

QY 87 WDOGNFPLIINKLIKIEDSDTYICEVEDQKEVQLVFGLTANS DTHLQGS LTLTLESP 146  
 DB 61 WDOGNFPLIINKLIKIEDSDTYICEVEDQKEVQLVFGLTANS DTHLQGS LTLTLESP 120

QY 147 PGSSPSVQCRSPRGKNIQSGKTLVSQLELDSDGTWCTVLOQKVEFKIDIVLAFQK 206  
 DB 121 PGSSPSVQCRSPRGKNIQSGKTLVSQLELDSDGTWCTVLOQKVEFKIDIVLAFQK 180

QY 207 ASSIYVKEGEQVEFSPLAFVTEKLTGSGELMWAERASSKSWITFDLKNKEVSVKRV 266  
 DB 181 ASSIYVKEGEQVEFSPLAFVTEKLTGSGELMWAERASSKSWITFDLKNKEVSVKRV 240

QY 267 TQDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKLHQEVNLVVMRATQLOK 326  
 DB 241 TQDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKLHQEVNLVVMRATQLOK 300

QY 327 TCEVWGPTSPKMLSLKLENKAKVSKREKPVWVLPNPEAGMOCCLSDSGVLLSNIKV 386  
 DB 301 TCEVWGPTSPKMLSLKLENKAKVSKREKPVWVLPNPEAGMOCCLSDSGVLLSNIKV 360

QY 387 LPTWSTPVHPRASALPAPPTGSALPDPTASALPDPRAASALPAAALAVISFLGLGIGV- 445  
 DB 361 LPTWSTPVHPRASALPAPPTGSALPDPTASALPDPRAASALPAAALAVISFLGLGIGV- 391

QY 446 ACVLARTR 453  
 DB 392 FCVRCRHR 399

RESULT 81

AAV30514  
 ID AAV30514 standard; protein; 432 AA.

AC AAV30514;

DT 15-NOV-1999 (first entry)

DE Predicted sequence of the CD4 protein.

XX CD4 protein; rapid mutational analysis method; protein epitope mapping;  
 XX binding domain mapping; binding capacity; anti-CD2 antibody;  
 XX anti-CD4 antibody; ligand binding site study.

OS Homo sapiens.

FN US955264-A.

PD 21-SEP-1999.

PE 11-OCT-1994; 94US-00320663.

XX 15-APR-1988; 88US-00181826.

PR 27-FEB-1992; 92US-00842465.

XX (GEHO ) GEN HOSPITAL CORP.

PI Seed B, Peterson A;

DR WPI; 1999-550602/46.

PT Rapid mutational analysis method for mapping protein epitopes.

PS Example 4; Fig 4; 27pp; English.

XX The present sequence represents the predicted sequence of the CD4

CC protein. The protein is used to demonstrate the method of the invention.

CC The specification describes a rapid mutational analysis method for

CC mapping protein epitopes and binding domains, by identifying substitution

CC mutations that result in the loss of binding capacity. The method may be

CC used for mapping protein epitopes, antigenic domains and binding sites.

CC It has been used for mapping binding sites for sixteen anti-CD2 and anti-

CC CD4 monoclonal antibodies. The method is especially useful for ligand

CC binding site studies for the design of new ligands and drugs

XX Sequence 432 AA;

Query Match 81.8%; Score 1901; DB 2; Length 432;

Best Local Similarity 88.6%; Pred. No. 1.3e-121;  
 Matches 379; Conservative 2; Mismatches 17; Indels 30; Gaps 2;

QY 27 KVLVGGKDPVLTCTTASQKSIQPHMKNSNOIKILGNQGSFLTKGPKLNDRADSRSL 86  
 DB 1 KVLVGGKDPVLTCTTASQKSIQPHMKNSNOIKILGNQGSFLTKGPKLNDRADSRSL 60

QY 87 WDOGNFPLIINKLIKIEDSDTYICEVEDQKEVQLVFGLTANS DTHLQGS LTLTLESP 146  
 DB 61 WDOGNFPLIINKLIKIEDSDTYICEVEDQKEVQLVFGLTANS DTHLQGS LTLTLESP 120

QY 147 PGSSPSVQCRSPRGKNIQSGKTLVSQLELDSDGTWCTVLOQKVEFKIDIVLAFQK 206  
 DB 121 PGSSPSVQCRSPRGKNIQSGKTLVSQLELDSDGTWCTVLOQKVEFKIDIVLAFQK 180

QY 207 ASSIYVKEGEQVEFSPLAFVTEKLTGSGELMWAERASSKSWITFDLKNKEVSVKRV 266  
 DB 181 ASSIYVKEGEQVEFSPLAFVTEKLTGSGELMWAERASSKSWITFDLKNKEVSVKRV 240

QY 267 TQDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKLHQEVNLVVMRATQLOK 326  
 DB 241 TQDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEKTKLHQEVNLVVMRATQLOK 300

QY 327 TCEVWGPTSPKMLSLKLENKAKVSKREKPVWVLPNPEAGMOCCLSDSGVLLSNIKV 386  
 DB 301 TCEVWGPTSPKMLSLKLENKAKVSKREKPVWVLPNPEAGMOCCLSDSGVLLSNIKV 360

QY 387 LPTWSTPVHPRASALPAPPTGSALPDPTASALPDPRAASALPAAALAVISFLGLGIGV- 445  
 DB 361 LPTWSTPVHPRASALPAPPTGSALPDPTASALPDPRAASALPAAALAVISFLGLGIGV- 391

QY 446 ACVLARTR 453  
 DB 392 FCVRCRHR 399

RESULT 82

ADA25188  
 ID ADA25188 standard; peptide; 432 AA.

AC ADA25188;

DT 20-NOV-2003 (first entry)

DE CD4 epitope region 1.

XX ligand binding site study; ligand design; drug design; CD4; epitope;

XX immunogenic.

```

OS Unidentified.
XX Key Location/Qualifiers
XX Region 34..52
FT Region /note="Leuza epitope"
FT Region 364..393
FT Region /note="Transmembrane region"
XX
XX US6579676-B1.
XX
XX 17-JUN-2003.
XX
XX 21-SEP-1999; 99US-000400207.
XX
XX 15-APR-1988; 88US-00181826.
XX 27-FEB-1992; 92US-00842465.
XX 11-OCT-1994; 94US-00320663.
XX
XX (GENO ) GEN HOSPITAL CORP.
XX
XX Seed B, Peterson A;
XX
XX WPI; 2003-644358/61.
XX
XX New population of recombinant cells collectively comprising mutant cDNAs,
XX useful in ligand binding site studies for the design of new ligands and
XX drugs.
XX
XX Disclosure; Fig 4; 27pp; English.
XX
XX The invention relates to a population of recombinant cells collectively
XX comprising mutant cDNAs comprising substitution mutants, where each
XX mutant cDNA encodes a particular protein of interest, where the
XX nucleotide sequence encoding the protein is known and the protein of
XX interest comprises at least two known binding domains comprising
XX different ligands. The population of recombinant cells collectively
XX comprising mutant cDNAs are useful in ligand binding site studies for the
XX design of new ligands and drugs. The present sequence represents the
XX amino acid sequence of the CD4 epitope region 1.
XX
XX Sequence 432 AA;
SQ
Query Match 81.8%; Score 1901; DB 6; Length 432;
Best Local Similarity 88.6%; Pred. No. 1.3e-121;
Matches 379; Conservative 2; Mismatches 17; Indels 30; Gaps 2;
QY 27 KYVLGKKGDVVELTCTASQKKSIOFHMKNMNOIKILGNOSFLTKGPSKLNDRASRRSL 86
DB 1 KYVLGKKGDVVELTCTASQKKSIOFHMKNMNOIKILGNOSFLTKGPSKLNDRASRRSL 60
QY 87 WDOGNFPLIIKNLKIDSDTYICEVEDQKEVQLVFGLTANSOTHLLOGSILTLTLESP 146
DB 61 WDOGNFPLIIKNLKIDSDTYICEVEDQKEVQLVFGLTANSOTHLLOGSILTLTLESP 120
QY 147 PGSSPEVOCRSPGKNIQGGKTLVSQLELQDSGTCTVLONOKKVERKIDIVLAFOK 206
DB 121 PGSSPEVOCRSPGKNIQGGKTLVSQLELQDSGTCTVLONOKKVERKIDIVLAFOK 180
QY 207 ASSIVYKKEGEQVEFSPFLAFVTEKLTSGGELMWQERASSSSKSWITPDLKNKEVSVKRV 266
DB 181 ASSIVYKKEGEQVEFSPFLAFVTEKLTSGGELMWQERASSSSKSWITPDLKNKEVSVKRV 240
QY 267 TODPKLQMGKULPLHLTLPQALPOVYAGSGLTALFAKTKGKHQEVNLVYMRATOLQKYL 326
DB 241 TODPKLQMGKULPLHLTLPQALPOVYAGSGLTALFAKTKGKHQEVNLVYMRATOLQKYL 300
QY 327 TCEVWGPPTSPKMLSLKLEKKEKAVSRKRPVWVNLPEAGMOCULSDSGOVLTLSSNIV 386
DB 301 TCEVWGPPTSPKMLSLKLEKKEKAVSRKRPVWVNLPEAGMOCULSDSGOVLTLSSNIV 360
QY 387 LPTWSTVHPRASALPAPPTGSALPDPTASALPDPPASALPALVALVSPFLGLGLGV- 445
DB 361 LPTWSTVHP-----MALIVLGVAIGLLPLFGLGLGF 391

```

```

QY 446 ACVLARTR 453
DB 392 FCVRCRHR 399
RESULT 83
AAP93528
ID AAP93528 standard; protein; 370 AA.
XX
XX AAP93528;
AC AAP93528;
DT 25-MAR-2003 (revised)
DT 04-JUN-1990 (first entry)
DE Human soluble CD4 protein (T4ex1) encoded by T4 SECL cDNA.
XX
XX Human soluble CD4 protein; T4 SECL cDNA; T4ex1;
XX HIV gp120 envelope protein; T-lymphocyte.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 129
FT Misc-difference 163
FT Misc-difference 183
FT Misc-difference 243
FT Misc-difference 370
XX
XX WO8903222-A.
XX
XX 20-APR-1989.
XX
XX 05-OCT-1988; 88WO-US003454.
XX
XX 08-OCT-1987; 87US-00106185.
XX 14-JAN-1988; 88US-00144313.
XX 14-JUN-1988; 88US-00206585.
XX 11-JUL-1988; 88US-00217475.
XX
XX (DANA-) DANA-FABER CANCER.
XX
XX Reinherz E, Husey R, Droski J, Richardson N;
PI WPI; 1989-129898/17.
XX
XX N-PSDB; AAN90764.
XX
XX Soluble human CD4 fragments - capable of binding HIV gp 120 envelope
XX protein, used for diagnosis, treatment or prophylaxis of HIV infection.
XX
XX Fig 1; pages 1/11-3/11; 106pp; English.
XX
XX Misc-difference in the features table represent where a modification in
XX cDNA and in the encoded CD4 protein has been made. A soluble human CD4
XX fragment capable of binding to the gp. 120 envelope protein of HIV is
XX claimed. Such a fragment should include none of the hydrophobic trans-
XX membrane region of naturally-occurring CD4, or contain a portion of the
XX hydrophobic region which is sufficiently short it does not prevent
XX solubilisation. Also claimed are modified soluble CD4 fragments which
XX differ by subseq. or addn. to this sequence (AAP93528) by at least one
XX AA. Such fragments pref. consist of: the first 182; 369; 549; 729 or 1107
XX AA; domain 1,2, and partial domain 3-pref. where the protein is
XX truncated at AA posn. 243; Asn-271 is Asp and/or Asn-300 is Asp; or the
XX AA sequence without the N-linked glycosylation sites, of CD4 protein.
XX Such fragments can also contain the corresp. AA of murine CD4 subseq. at
XX more than one triplet site in the DNA, selected from a triplet encoding
XX the AA at posn. 48; 50; 51; 121; 122; 123; 155; 156; or 158 of human CD4.
XX CD4 fragments have the capacity to prevent infection of human T-
XX lymphocyte infection by HIV and to prevent the formation of human T-
XX lymphocyte syncytia. They can be used in the diagnosis, treatment and
XX prevention of HIV infection. (Updated on 25-MAR-2003 to correct PR
XX field.) (Updated on 25-MAR-2003 to correct PI field.)
XX

```

SQ Sequence 370 AA;

Query Match 81.7%; Score 1899; DB 1; Length 370;  
Best Local Similarity 99.7%; Pred. No. 1.4e-121;  
Matches 368; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 27 KVLGKKGDVVELCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGSPSKLNDRADSRSL 86  
DB 2 KVLGKKGDVVELCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGSPSKLNDRADSRSL 61  
QY 87 WDQGNFPLIINKLIKIEDSDTYICEVEDQKEEVQLLVFGLTANSPTHLLQGSLLTLESP 146  
DB 62 WDQGNFPLIINKLIKIEDSDTYICEVEDQKEEVQLLVFGLTANSPTHLLQGSLLTLESP 121  
QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWCTVQONQKVEFKIDIVLAFQK 206  
DB 122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWCTVQONQKVEFKIDIVLAFQK 181  
QY 207 ASSIYVKEGEQVEFSFPLAFVEKLTGSGELMWQAEARASSSKSWITFDLKNKEVSVKRV 266  
DB 182 ASSIYVKEGEQVEFSFPLAFVEKLTGSGELMWQAEARASSSKSWITFDLKNKEVSVKRV 241  
QY 267 TQDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEATGKLDQEVNLVVMRATQLOKNL 326  
DB 242 TQDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEATGKLDQEVNLVVMRATQLOKNL 301  
QY 327 TCEVWGPTSPKLMSTLKENKEAKVSKREKPVWVLNPEAGMOCCLSDSGOVLLESNIKV 386  
DB 302 TCEVWGPTSPKLMSTLKENKEAKVSKREKPVWVLNPEAGMOCCLSDSGOVLLESNIKV 361  
QY 387 LPTWSTPVH 395  
DB 362 LPTWSTPVH 370

RESULT 84

AAB83356  
ID AAB83356 standard; protein; 370 AA.

AC AAB83356;  
XX  
DT 09-OCT-2001 (first entry)  
XX  
DE Human CD4 protein sequence.  
XX  
KW Chemotactic chemokine receptor 5; gp120; CD4; therapy; HIV; CCR5;  
KW human immunodeficiency virus; anti-inflammatory disease; human.  
XX  
OS Homo sapiens.  
XX  
PN EP118858-A2.  
XX  
PD 25-JUL-2001.  
XX  
PF 03-JAN-2001; 2001EP-0030020.  
XX  
PR 12-JAN-2000; 2000GB-00000659.  
PR 12-JAN-2000; 2000GB-00000661.  
PR 12-JAN-2000; 2000GB-00000663.  
XX  
PA (PF12 ) PFIZER LTD.  
PA (PF12 ) PFIZER INC.  
XX  
PI Dobbs S, Petros M, Rickett GA;  
XX  
DR WP1; 2001-477088/52.  
DR N-PSDB; AAF87101.  
XX  
PT Determining if an agent can modulate CCR5-gp120 interaction. comprises  
PT incubating the agent with CCR5 and gp120 and determining if the agent  
PT modulates the interaction.  
XX  
PS Claim 1; Page 112; 113PD; English.

XX This sequence represents the human CD4 protein sequence. The invention  
CC relates to a method for determining whether an agent is capable of  
CC modulating the interaction of chemotactic chemokine receptor 5 (CCR5)  
CC with gp120, comprising incubating the agent with CCR5 and gp120 and  
CC determining whether the agent modulates the interaction, where gp120 is  
CC associated with CD4, and where the interaction is a low affinity binding.  
CC The method is used to identify an agent capable of modulating the  
CC interaction of CCR5 with gp120. An agent identified by the method is used  
CC to prepare a pharmaceutical composition for the treatment of a disease or  
CC condition associated with CCR5 and gp120 interaction, to treat a subject  
CC with a disease or condition associated with CCR5 and gp120 interaction,  
CC and for preparing a pharmaceutical for treating human immunodeficiency  
CC virus (HIV). It can also be used to treat anti-inflammatory diseases. The method  
CC is commercially useful, amenable to high throughput screening, and  
CC detects interaction of gp120 with cells expressing only CCR5  
XX

SQ Sequence 370 AA;

Query Match 81.7%; Score 1899; DB 4; Length 370;  
Best Local Similarity 99.7%; Pred. No. 1.4e-121;  
Matches 368; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 27 KVLGKKGDVVELCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGSPSKLNDRADSRSL 86  
DB 2 KVLGKKGDVVELCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGSPSKLNDRADSRSL 61  
QY 87 WDQGNFPLIINKLIKIEDSDTYICEVEDQKEEVQLLVFGLTANSPTHLLQGSLLTLESP 146  
DB 62 WDQGNFPLIINKLIKIEDSDTYICEVEDQKEEVQLLVFGLTANSPTHLLQGSLLTLESP 121  
QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWCTVQONQKVEFKIDIVLAFQK 206  
DB 122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWCTVQONQKVEFKIDIVLAFQK 181  
QY 207 ASSIYVKEGEQVEFSFPLAFVEKLTGSGELMWQAEARASSSKSWITFDLKNKEVSVKRV 266  
DB 182 ASSIYVKEGEQVEFSFPLAFVEKLTGSGELMWQAEARASSSKSWITFDLKNKEVSVKRV 241  
QY 267 TQDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEATGKLDQEVNLVVMRATQLOKNL 326  
DB 242 TQDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALAEATGKLDQEVNLVVMRATQLOKNL 301  
QY 327 TCEVWGPTSPKLMSTLKENKEAKVSKREKPVWVLNPEAGMOCCLSDSGOVLLESNIKV 386  
DB 302 TCEVWGPTSPKLMSTLKENKEAKVSKREKPVWVLNPEAGMOCCLSDSGOVLLESNIKV 361  
QY 387 LPTWSTPVH 395  
DB 362 LPTWSTPVH 370

RESULT 85

AA39824  
ID AA39824 standard; protein; 369 AA.

AC AA39824;  
XX  
DT 03-DEC-1999 (first entry)  
XX  
DE Soluble human T4 protein.  
XX  
KW Soluble T4 protein; gT4; human; HIV; binding inhibitor; T4+ cell; AIDS;  
KW vaccine; immunisation; therapy.  
XX  
OS Homo sapiens.  
XX  
PN US5958678-A.  
XX  
PD 28-SEP-1999.  
XX  
PF 12-DEC-1994; 94US-00354452.  
XX

```
PR 21-AUG-1986; 86US-00898587.  
PR 11-JUN-1991; 91US-00713564.  
PR 06-JUL-1992; 92US-00908021.  
XX  
XX (UYCO ) UNIV COLUMBIA NEW YORK.  
XX PA  
XX PI McDougal JS, Weiss R, Axel R, Littman DR, Madden PJ, Chess L;  
XX DR WPI; 1999-561025/47.  
XX  
XX Human T4 protein inhibits HIV binding to T4 cells, useful for treating  
XX PT AIDS.  
XX PS  
XX Claim 1; Col 51-53; 58pp; English.  
XX  
CC This sequence represents the soluble human T4 protein of the invention.  
CC The soluble human T4 protein blocks the binding of HIV to T4+ cells and  
CC is therefore useful for the treatment of AIDS. Monoclonal antibodies  
CC against the T4 protein may be used as vaccines for immunizing subjects  
CC against AIDS  
CC  
XX Sequence 369 AA;  
SQ  
Query Match 81.3%; Score 1891; DB 2; Length 369;  
Best Local Similarity 99.7%; Pred. No. 5,1e-121;  
Matches 367; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 27 KVLVAGKGDVVELTCTASQKSIQPHWKNNSQIKILNGSGFLLTKGPKLNDRADSRSL 86  
DB 2 KVLVAGKGDVVELTCTASQKSIQPHWKNNSQIKILNGSGFLLTKGPKLNDRADSRSL 61  
QY 87 WDQGNFPLIIKRLKIEDSDTYICEVEDQKEVQLVFGLTANSPTHLQGGSLTTLTLESP 146  
DB 62 WDQGNFPLIIKRLKIEDSDTYICEVEDQKEVQLVFGLTANSPTHLQGGSLTTLTLESP 121  
QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONOKVEFKIDIVVLAPOK 206  
DB 122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONOKVEFKIDIVVLAPOK 181  
QY 207 ASSIYKKEGGEVSEFPLAFTVEKLTGSGELMWQAERASSSKSWITFDLNKKEVSVKRV 266  
DB 182 ASSIYKKEGGEVSEFPLAFTVEKLTGSGELMWQAERASSSKSWITFDLNKKEVSVKRV 241  
QY 267 TDDPKLQWKGKLPRLHLLTPQALPOYAGSGNTLLALEAKTGKLDHGVNLVWRAATOLQKNL 326  
DB 242 TDDPKLQWKGKLPRLHLLTPQALPOYAGSGNTLLALEAKTGKLDHGVNLVWRAATOLQKNL 301  
QY 327 TCEVWGPTSPKLMLSLKLENKAKVSKREKPYWVLNPEAGMOCILSDSGOVLLESNTKV 386  
DB 302 TCEVWGPTSPKLMLSLKLENKAKVSKREKPYWVLNPEAGMOCILSDSGOVLLESNTKV 361  
QY 387 LPTWSTPV 394  
DB 362 LPTWSTPV 369  
RESULT 86  
AA88327 standard; protein; 369 AA.  
ID AAY88327;  
XX  
XX AAY88327;  
XX  
XX 14-JUL-2000 (first entry)  
XX  
XX T4 glycoprotein amino acid sequence.  
XX DE  
XX T4 glycoprotein amino acid sequence.  
XX  
XX sT4; glycoprotein; human immunodeficiency virus; HIV; block binding;  
XX KM AIDS; treatment; inhibits; cell to cell spread; infection; fusion.  
XX  
XX Mammalia.  
XX OS  
XX  
XX US5126433-A.  
XX
```

```
PD 30-JUN-1992.  
XX  
XX 23-OCT-1987; 87US-00114244.  
XX  
XX PR 21-AUG-1986; 86US-00898587.  
XX  
XX (UYCO ) UNIV COLUMBIA NEW YORK.  
XX PA  
XX PI Madden PJ, Chess L, Axel R, Weiss R, Littman DR, McDougal JS;  
XX DR WPI; 2000-348913/30.  
XX  
XX  
XX Soluble T4 glycoprotein useful for prevention and treatment of acquired  
XX PT immunodeficiency syndrome and for screening inhibitors of human  
XX PT immunodeficiency viral binding.  
XX  
XX Claim 1; Col 54; 64pp; English.  
XX  
XX This sequence represents the amino acid sequence of glycosylated sT4  
XX glycoprotein. Human immunodeficiency virus (HIV) uses sT4 as a target  
XX receptor on T cells. The invention relates to glycosylated sT4 which  
XX can be used for the prophylaxis and treatment of AIDS patients. Administration  
XX of sT4 effectively inhibits the cell to cell spreading of HIV infection  
XX and also the fusion of HIV-infected T4 cells and non-infected T4 cells.  
XX The administration of T4 alleviates several symptoms associated with  
XX CC AIDS, and prevents the occurrence of new pathological changes. The sT4  
XX glycoprotein is useful for the prophylaxis and treatment of patients with  
XX CC AIDS. It is also useful as a reagent to identify natural, synthetic or  
XX recombinant molecules which act as therapeutic agents or inhibitors of  
XX T4+ cell interactions and in diagnostic assays for detection T4 proteins  
XX or molecules  
XX  
XX Sequence 369 AA;  
SQ  
Query Match 81.3%; Score 1891; DB 3; Length 369;  
Best Local Similarity 99.7%; Pred. No. 5,1e-121;  
Matches 367; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 27 KVLVAGKGDVVELTCTASQKSIQPHWKNNSQIKILNGSGFLLTKGPKLNDRADSRSL 86  
DB 2 KVLVAGKGDVVELTCTASQKSIQPHWKNNSQIKILNGSGFLLTKGPKLNDRADSRSL 61  
QY 87 WDQGNFPLIIKRLKIEDSDTYICEVEDQKEVQLVFGLTANSPTHLQGGSLTTLTLESP 146  
DB 62 WDQGNFPLIIKRLKIEDSDTYICEVEDQKEVQLVFGLTANSPTHLQGGSLTTLTLESP 121  
QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONOKVEFKIDIVVLAPOK 206  
DB 122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONOKVEFKIDIVVLAPOK 181  
QY 207 ASSIYKKEGGEVSEFPLAFTVEKLTGSGELMWQAERASSSKSWITFDLNKKEVSVKRV 266  
DB 182 ASSIYKKEGGEVSEFPLAFTVEKLTGSGELMWQAERASSSKSWITFDLNKKEVSVKRV 241  
QY 267 TDDPKLQWKGKLPRLHLLTPQALPOYAGSGNTLLALEAKTGKLDHGVNLVWRAATOLQKNL 326  
DB 242 TDDPKLQWKGKLPRLHLLTPQALPOYAGSGNTLLALEAKTGKLDHGVNLVWRAATOLQKNL 301  
QY 327 TCEVWGPTSPKLMLSLKLENKAKVSKREKPYWVLNPEAGMOCILSDSGOVLLESNTKV 386  
DB 302 TCEVWGPTSPKLMLSLKLENKAKVSKREKPYWVLNPEAGMOCILSDSGOVLLESNTKV 361  
QY 387 LPTWSTPV 394  
DB 362 LPTWSTPV 369  
RESULT 87  
AAR12956 standard; protein; 370 AA.  
ID AAR12956  
XX  
XX AAR12956;  
XX
```

```

XX 25-MAR-2003 (revised)
DT 09-JAN-2003 (revised)
DT 18-SEP-1991 (first entry)
XX
DE Lys(64) CD4 mutant.
XX
XX CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;
XX human immunodeficiency virus.
XX
XX Homo sapiens.
XX
XX WO9109123-A.
XX
XX 27-JUN-1991.
XX
XX 15-DEC-1989; 89WO-US005625.
XX
XX 15-DEC-1989; 89WO-US005625.
XX
XX (DAND ) DANA FARBER CANCER INST INC.
XX (REIN/) REINHERZ E L.
XX
XX Reinherz E, Clayton LK;
XX
XX WPI; 1991-208149/28.
XX
XX New soluble human CD4 glyco:protein - has decreased affinity for MHC
XX class II antigens and is used to treat, prevent and diagnose HIV
XX infections.
XX
XX Disclosure; Table 1; 41pp; English.
XX
XX Mutant M4 is one of 17 modified CD4 proteins analysed for class II MHC
XX binding. It corresponds to domains I and II of the human CD4 protein in
XX which Gln at position 64 is replaced with Lys. Mutant M4 is not one of
XX the preferred mutants of the invention as it retains its Class II MHC
XX binding affinity. See AARI2951-R12967. (Updated on 09-JAN-2003 to add
XX missing OS field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated
XX on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 370 AA;
SQ
Query Match 81.2%; Score 1887; DB 2; Length 370;
Best Local Similarity 99.2%; Pred. No. 9.6e-121;
Matches 366; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 27 KVLVGGKGDVVELTCTASQKKSIOFHMKNSNOIKILGNQGSFLTKGPSKLNDRADSRSL 86
DB 2 KVLVGGKGDVVELTCTASQKKSIOFHMKNSNOIKILGNQGSFLTKGPSKLNDRADSRSL 61
QY WDCGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLQOGSLTLTLESP 146
DB 62 WDCGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLQOGSLTLTLESP 121
QY 147 PGSSPSVQCSRPRGKNIQGGKTLVSQLELDOSGTWTCVYLQNKVEFKIDIVLAFOK 206
DB 122 PGSSPSVQCSRPRGKNIQGGKTLVSQLELDOSGTWTCVYLQNKVEFKIDIVLAFOK 181
QY 207 ASSIVYKKEGEQVEFSPPLAFTVEKLTGSGELMWQAERASSSKSWITFDLKNKESYKRV 266
DB 182 ASSIVYKKEGEQVEFSPPLAFTVEKLTGSGELMWQAERASSSKSWITFDLKNKESYKRV 241
QY 267 TODPKLQMGKPLHLTLPOALPOYAGSGNLTALAEATGKLHDEVNLVVMRAOTOLQKNL 326
DB 242 TODPKLQMGKPLHLTLPOALPOYAGSGNLTALAEATGKLHDEVNLVVMRAOTOLQKNL 301
QY 327 TCEVWGPTSPKLMNLSLKLENKEAKVSKREKPVWVLPNPAQMWQCLSDSGVLLSNIKV 386
DB 302 TCEVWGPTSPKLMNLSLKLENKEAKVSKREKPVWVLPNPAQMWQCLSDSGVLLSNIKV 361
QY 387 LPTWSTPVH 395
|||||

```

```

DB 362 LPTWSTPVH 370
RESULT 88
AARI2954
ID AARI2954 standard; protein; 370 AA.
XX
XX AARI2954;
XX
XX 25-MAR-2003 (revised)
DT 09-JAN-2003 (revised)
DT 18-SEP-1991 (first entry)
XX
XX His(40) CD4 mutant.
XX
XX CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;
XX human immunodeficiency virus.
XX
XX Homo sapiens.
XX
XX WO9109123-A.
XX
XX 27-JUN-1991.
XX
XX 15-DEC-1989; 89WO-US005625.
XX
XX 15-DEC-1989; 89WO-US005625.
XX
XX (DAND ) DANA FARBER CANCER INST INC.
XX (REIN/) REINHERZ E L.
XX
XX Reinherz E, Clayton LK;
XX
XX WPI; 1991-208149/28.
XX
XX New soluble human CD4 glyco:protein - has decreased affinity for MHC
XX class II antigens and is used to treat, prevent and diagnose HIV
XX infections.
XX
XX Disclosure; Table 1; 41pp; English.
XX
XX Mutant M2 is one of 17 modified CD4 proteins analysed for class II MHC
XX binding. It corresponds to domains I and II of the human CD4 protein in
XX which Gln at position 40 is replaced with His. Mutant M2 is not one of
XX the preferred mutants of the invention as it retains its Class II MHC
XX binding affinity. See AARI2951-R12967. (Updated on 09-JAN-2003 to add
XX missing OS field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated
XX on 25-MAR-2003 to correct PA field.)
XX
XX Sequence 370 AA;
SQ
Query Match 81.1%; Score 1886; DB 2; Length 370;
Best Local Similarity 99.2%; Pred. No. 1.1e-120;
Matches 366; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 27 KVLVGGKGDVVELTCTASQKKSIOFHMKNSNOIKILGNQGSFLTKGPSKLNDRADSRSL 86
DB 2 KVLVGGKGDVVELTCTASQKKSIOFHMKNSNOIKILGNQGSFLTKGPSKLNDRADSRSL 61
QY WDCGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLQOGSLTLTLESP 146
DB 62 WDCGNFPLIIKNLKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLQOGSLTLTLESP 121
QY 147 PGSSPSVQCSRPRGKNIQGGKTLVSQLELDOSGTWTCVYLQNKVEFKIDIVLAFOK 206
DB 122 PGSSPSVQCSRPRGKNIQGGKTLVSQLELDOSGTWTCVYLQNKVEFKIDIVLAFOK 181
QY 207 ASSIVYKKEGEQVEFSPPLAFTVEKLTGSGELMWQAERASSSKSWITFDLKNKESYKRV 266
DB 182 ASSIVYKKEGEQVEFSPPLAFTVEKLTGSGELMWQAERASSSKSWITFDLKNKESYKRV 241
QY 267 TODPKLQMGKPLHLTLPOALPOYAGSGNLTALAEATGKLHDEVNLVVMRAOTOLQKNL 326
|||||

```



Dd		2422 TQDPTLQWGMKGLPRLHITLPQLPVYAGSGNLTALAEAKTQKGLHQEVNLVVMRATQIQKNL	361
Oy		327 TCENVGPTSPKMLSLIKLENKEAKSKREKXPVAVLNPEAGMOCCLSDSQVILLESNIKV	386
Dd		302 TCENVGPTSPKMLSLIKLENKEAKSKREKXAVVLNPEAGMOCCLSDSQVILLESNIKV	361
Oy		367 LPTWSTPPVH 395 	
Dd		362 LPTWSTPPVH 370	
		RESULT 89	
	ID	AAR12964 standard; protein; 370 AA.	
	XX	AAR12964;	
	AC		
	XX		
	DT	25-MAR-2003 (revised)	
	DT	09-JAN-2003 (revised)	
	DT	18-SEP-1991 (first entry)	
	DE	Val(143) CD4 mutant.	
	XX		
	KM	CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;	
	KW	human immunodeficiency virus.	
	OS	Homo sapiens.	
	PN	MOJ019123-A.	
	PD	27-JUN-1991.	
	XX		
	PF	15-DEC-1989; 89WO-US005625.	
	XX		
	PR	15-DEC-1989; 89WO-US005625.	
	XX		
	PA	(DAND ) DANA PARBER CANCER INST INC. (REIN/) REINHHERZ E L.	
	PI	Reinherz E, Clayton LK;	
	DR	WPI; 1991-208149/28.	
	XX		
	PT	New soluble human CD4 glyco:protein - has decreased affinity for MHC	
	PT	class II antigens and is used to treat, prevent and diagnose HIV	
	PS	infections.	
	XX		
	PS	Disclosure; Table 1; 41pp; English.	
	CC	Mutant M12 is one of 17 modified CD4 proteins analysed for class II MHC	
	CC	binding. It corresponds to domains I and II of the human CD4 protein in	
	CC	which Thr at position 143 is replaced with Val. Mutant M12 is not one of	
	CC	the preferred mutants of the invention as it retains its Class II MHC	
	CC	binding affinity. See AAR12951-R12967. (updated on 09-JAN-2003 to add	
	CC	missing OS field.) (updated on 25-MAR-2003 to correct PR field.) (updated	
	CC	on 25-MAR-2003 to correct PA field.)	
	XX		
SQ		Sequence 370 AA;	
		Query Match 81.1%; Score 1886; DB 2; Length 370;	
		Best Local Similarity 99.2%; Pred. No. 1.1e-120;	
		Matches 366; Conservative 0; Mismatches 3; Indels 0; Gaps 0	
Oy		27 KVVLGKKGDVTVELTCASQKSIOFHMKNSNOTKITGNQOSFLTCKGPSKLNDPADSRSL	86
Dd		2 KVVLGKKGDVTVELTCASQKSIQFMKNSNOJKITGNQOSFLTCKGPSKLNDPADSRSL	61
Oy		87 WDQGFPLIIIKVLKIEDSDTYICEVEDQKEEVOLVFGLTANSDFHLQGOSTLTLLSEF	146
Dd		62 WDOGFPFLIIRKVLKIEDSDTYICEVEDQKEEVOLVFGLTANSDFHLQGOSTLTLLSEF	121
Oy		147 PGSSFSVQGRSRRGKNIQGGKTLVSQLELDSDGTTCITYLNQKKVEFKIDIVLAFOK	206

Db	122	PSSPSVQCRSPRGKNIIGGKVLVSVDLELDSGWCTTVALQNKKVEFKDIDVLAFOK	186
Qy	207	ASSIVYKKEGEQVESFFPLAFVTEKLTSGGELMWAQEPASSSKWITFDLKNNKESVKRV	266
Db	182	ASSIVYKKEGEQVESFFPLAFVTEKLTSGGELMWAQEPASSSKWITFDLKNNKESVKRV	241
Qy	267	TQDPKIQMGKGLPHLTLPQALPOVAGSGNLTLMLEAKTGLDHENVLVVMRATOLQNL	326
Db	242	TQDPKIQMGKGLPHLTLPQALPOVAGSGNLTLMLEAKTGLDHENVLVVMRATOLQNL	301
Qy	327	TCEVMGPTSPKMLSLIKLENKAQYSKREKPVMVILNPAGMMOCLLSDSQVLLSNIKV	386
Db	302	TCEVMGPTSPKMLSLIKLENKAQYSKREKAVMVILNPAGMMOCLLSDSQVLLSNIKV	361
Qy	387	LPTWSTPVIH 395       	
Db	362	LPTWSTPVIH 370       	
RESULT 90			
ID	AARI2965	AARI2965 standard; protein; 370 AA.	
XX	AC	AARI2965;	
XX	DT	25-MAR-2003 (revised)	
XX	DT	09-JAN-2003 (revised)	
XX	DT	18-SEP-1991 (first entry)	
XX	DE	Arg(150) CD4 mutant.	
XX	KW	CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS; human immunodeficiency virus.	
XX	KS	Homo sapiens.	
XX	PN	M09109123-A.	
XX	PD	27-JUN-1991.	
XX	PF	15-DEC-1989; 89MO-US005625.	
XX	PR	15-DEC-1989; 89MO-US005625.	
XX	PA	(DAND ) DANA FARBER CANCER INST INC. (REIN/) REINHERTZ E L.	
XX	PI	Reinherz E, Clayton LK;	
XX	DR	WPI; 1991-208149/28.	
XX	PT	New soluble human CD4 glyco:protein - has decreased affinity for MHC Class II antigens and is used to treat, prevent and diagnose HIV infections.	
XX	PS	Disclosure; Table 1; 41pp; English.	
XX	CC	Mutant M13 is one of 17 modified CD4 proteins analysed for class II MHC binding. It corresponds to domains I and II of the human CD4 protein in which Gln at position 150 is replaced with Arg. Mutant M13 is not one of the preferred mutants of the invention as it retains its Class II MHC binding affinity. See AARI2951-R12967. (Updated on 09-JAN-2003 to add missing OS field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to correct PA field.)	
XX	SO	Sequence 370 AA;	
Query Match                      81.1%; Score 1886; DB 2; Length 370;			
Best Local Similarity    99.2%; Pred.No. 1,le-120;			
Matches    366; Conservative    0; Mismatches    3; Indels    0; Gaps    0			
Qy	27	KVYLKGKGDVIELTCASQKSKIQHWMKNSNQIKLGNQSFLTKGSGKLNDRADSRRL	86

Db 2 KVLGKKGDVVELTCTASQKSIQFHWKNSNQIKLNGQSFLLTKGPSKLNDRADSRSL 61  
 QY 87 WDOGNFPLIIKNLKIEDSDTYICEVEDQKEVOLLVFGLTANSDTHLLQGOSLTLTLESP 146  
 Db 62 WDOGNFPLIIKNLKIEDSDTYICEVEDQKEVOLLVFGLTANSDTHLLQGOSLTLTLESP 121  
 QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLOQOKVEFKIDIVLAFQK 206  
 Db 122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLOQOKVEFKIDIVLAFQK 181  
 QY 207 ASSIVYKKEGQVEFSFPLAFIVEKLTGSGELMWQAERASSSKSWITFDLKNKESVYKRV 266  
 Db 182 ASSIVYKKEGQVEFSFPLAFIVEKLTGSGELMWQAERASSSKSWITFDLKNKESVYKRV 241  
 QY 267 TQDPKLOMGKKLPPLHLTLPQALPOYAGSGNLTALAEKTKLHOEVNLVVMRATQLOKNL 326  
 Db 242 TQDPKLOMGKKLPPLHLTLPQALPOYAGSGNLTALAEKTKLHOEVNLVVMRATQLOKNL 301  
 QY 327 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVVNLNPEAGMOCCLSDSGVLLLESNIKV 386  
 Db 302 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVVNLNPEAGMOCCLSDSGVLLLESNIKV 361  
 QY 387 LPTWSTPVH 395  
 Db 362 LPTWSTPVH 370

## RESULT 91

AAR12958  
 ID AAR12958 standard; protein; 370 AA.

AC AAR12958;

XX 25-MAR-2003 (revised)  
 DT 09-JAN-2003 (revised)  
 DT 18-SEP-1991 (first entry)

XX Gln(80) CD4 mutant.

XX CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;  
 KW human immunodeficiency virus.

XX Homo sapiens.

XX WO9109123-A.

XX 27-JUN-1991.

XX 15-DEC-1989; 89WO-US005625.

XX 15-DEC-1989; 89WO-US005625.

PA (DAND ) DANA FARBER CANCER INST INC.  
 PA (REIN/) REINHERZ E L.

XX Reinherz E, Clayton LK;

XX WPI; 1991-208149/28.

XX New soluble human CD4 glyco:protein - has decreased affinity for MHC  
 PT class II antigens and is used to treat, prevent and diagnose HIV  
 PT infections.

XX Disclousure; Table 1; 41pp; English.

XX Mutant M6 is one of 17 modified CD4 proteins analysed for class II MHC  
 CC binding. It corresponds to domains I and II of the human CD4 protein in  
 CC which Asp at position 80 is replaced with Gln. Mutant M6 is not one of  
 CC the preferred mutants of the invention as it could not be evaluated for  
 CC class II MHC binding affinity. The substitutions grossly affect the  
 CC structure of the external CD4 domains so that reactivity with all anti-  
 CC CD4 monoclonal antibodies tested was reduced. See AAR12951-R12967.  
 CC (Updated on 09-JUN-2003 to add missing OS field.) (Updated on 25-MAR-2003

CC to correct PR field.) (Updated on 25-MAR-2003 to correct PA field.)  
 XX Sequence 370 AA;

Query Match 81.1%; Score 1885; DB 2; Length 370;  
 Best Local Similarity 99.2%; Pred. No. 1.3e-120;  
 Matches 366; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 27 KVLGKKGDVVELTCTASQKSIQFHWKNSNQIKLNGQSFLLTKGPSKLNDRADSRSL 86  
 Db 2 KVLGKKGDVVELTCTASQKSIQFHWKNSNQIKLNGQSFLLTKGPSKLNDRADSRSL 61  
 QY 87 WDOGNFPLIIKNLKIEDSDTYICEVEDQKEVOLLVFGLTANSDTHLLQGOSLTLTLESP 146  
 Db 62 WDOGNFPLIIKNLKIEDSDTYICEVEDQKEVOLLVFGLTANSDTHLLQGOSLTLTLESP 121  
 QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLOQOKVEFKIDIVLAFQK 206  
 Db 122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLOQOKVEFKIDIVLAFQK 181  
 QY 207 ASSIVYKKEGQVEFSFPLAFIVEKLTGSGELMWQAERASSSKSWITFDLKNKESVYKRV 266  
 Db 182 ASSIVYKKEGQVEFSFPLAFIVEKLTGSGELMWQAERASSSKSWITFDLKNKESVYKRV 241  
 QY 267 TQDPKLOMGKKLPPLHLTLPQALPOYAGSGNLTALAEKTKLHOEVNLVVMRATQLOKNL 326  
 Db 242 TQDPKLOMGKKLPPLHLTLPQALPOYAGSGNLTALAEKTKLHOEVNLVVMRATQLOKNL 301  
 QY 327 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVVNLNPEAGMOCCLSDSGVLLLESNIKV 386  
 Db 302 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVVNLNPEAGMOCCLSDSGVLLLESNIKV 361  
 QY 387 LPTWSTPVH 395  
 Db 362 LPTWSTPVH 370

## RESULT 92

AAR12951  
 ID AAR12951 standard; protein; 370 AA.

AC AAR12951;

XX 25-MAR-2003 (revised)  
 DT 09-JAN-2003 (revised)  
 DT 18-SEP-1991 (first entry)

XX Gln(17), Ser(18) CD4 mutant.

XX CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;  
 KW human immunodeficiency virus.

XX Homo sapiens.

XX WO9109123-A.

XX 27-JUN-1991.

XX 15-DEC-1989; 89WO-US005625.

XX 15-DEC-1989; 89WO-US005625.

PA (DAND ) DANA FARBER CANCER INST INC.  
 PA (REIN/) REINHERZ E L.

XX Reinherz E, Clayton LK;

XX WPI; 1991-208149/28.

XX New soluble human CD4 glyco:protein - has decreased affinity for MHC  
 PT class II antigens and is used to treat, prevent and diagnose HIV  
 PT infections.

PS Claim 3; Table 1; 41pp; English.

XX Mutant M1.1 is one example of a novel CD4 protein of the invention. It  
 CC corresponds to domains I and II of the human CD4 protein in which Thr at  
 CC position 17 and Ala at position 18 are replaced with Glu and Ser,  
 CC respectively. The modified CD4 can inhibit HIV binding to the CD4  
 CC receptors of T lymphocytes and deplete circulating levels of HIV gp120;  
 CC it does not bind class II antigen bearing B lymphocytes. See also  
 CC AAR12952-R12967. (Updated on 03-JAN-2003 to add missing OS field.)  
 CC (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to  
 CC correct PA field.)  
 CC  
 XX Sequence 370 AA;  
 SQ

Query Match 80.9%; Score 1882; DB 2; Length 370;  
 Best Local Similarity 98.9%; Pred. No. 2.1e-120;  
 Matches 365; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 27 KVLGKKGDVVELCTASQKSIQFHWKNSNQIKILNGSGFLLTKGPSKLANDRADSRSL 86  
 DB 2 KVLGKKGDVVELCTESSQKSIQFHWKNSNQIKILNGSGFLLTKGPSKLANDRADSRSL 61

QY 87 WDQGNFPLIIKMLKIEDSDTYICEVEDQKEVQLVFGLTANSDTHLQGGSLTLTLESP 146  
 DB 62 WDQGNFPLIIKMLKIEDSDTYICEVEDQKEVQLVFGLTANSDTHLQGGSLTLTLESP 121

QY 147 PSSSSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONQKVEFKIDIVLAFQK 206  
 DB 122 PSSSSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONQKVEFKIDIVLAFQK 181

QY 207 ASSIYKKEGEQVEFSFLAFTVEKLTGSGELMWQAEASSSKSWITFDLKNKEYSVKRV 266  
 DB 182 ASSIYKKEGEQVEFSFLAFTVEKLTGSGELMWQAEASSSKSWITFDLKNKEYSVKRV 241

QY 267 TQDPKLQMGKKLPLHLTLPOALPOYAGSGNLTALAEAKTGKLEHVNLVVWRATOLQKNL 326  
 DB 242 TQDPKLQMGKKLPLHLTLPOALPOYAGSGNLTALAEAKTGKLEHVNLVVWRATOLQKNL 301

QY 327 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLSDSGVLLLESNTKV 386  
 DB 302 TCEVWGPTSPKMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLSDSGVLLLESNTKV 361

QY 387 LPTWSTPVH 395  
 DB 362 LPTWSTPVH 370

RESULT 93  
 AAR12957  
 ID AAR12957 standard; protein; 370 AA.  
 XX  
 AC AAR12957;  
 XX  
 XX 25-MAR-2003 (revised)  
 DT 09-JAN-2003 (revised)  
 DT 18-SEP-1991 (first entry)  
 XX  
 DE Asn(72), Lys(73) CD4 mutant.  
 XX  
 KM CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;  
 KW human immunodeficiency virus.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9109123-A.  
 XX  
 PD 27-JUN-1991.  
 XX  
 XX 15-DEC-1989; 89WO-US005625.  
 PF  
 XX 15-DEC-1989; 89WO-US005625.  
 PR  
 XX 15-DEC-1989; 89WO-US005625.  
 XX  
 PA (DAND) DANA FARBER CANCER INST INC.

PA (REIN/) REINHERZ E L.  
 XX  
 PI Reinherz E, Clayton UK;  
 XX  
 DR WPI, 1991-208149/28.  
 XX  
 PT New soluble human CD4 glyco:protein - has decreased affinity for MHC  
 PT class II antigens and is used to treat, prevent and diagnose HIV  
 PT infections.  
 XX  
 XX Claim 3; Table 1; 41pp; English.  
 PS  
 CC Mutant M5 is one example of a novel CD4 protein of the invention. It  
 CC corresponds to domains I and II of the human CD4 protein in which Lys at  
 CC position 72 and Asn at position 73 are replaced with Asn and Lys,  
 CC respectively. The modified CD4 can inhibit HIV binding to the CD4  
 CC receptors of T lymphocytes and deplete circulating levels of HIV gp120;  
 CC it does not bind class II antigen bearing B lymphocytes. See AAR12951-  
 CC R12967. (Updated on 09-JAN-2003 to add missing OS field.) (Updated on 25-  
 CC MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to correct PA  
 CC field.)  
 CC  
 XX Sequence 370 AA;  
 SQ

Query Match 80.9%; Score 1880; DB 2; Length 370;  
 Best Local Similarity 98.9%; Pred. No. 2.9e-120;  
 Matches 365; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 27 KVLGKKGDVVELCTASQKSIQFHWKNSNQIKILNGSGFLLTKGPSKLANDRADSRSL 86  
 DB 2 KVLGKKGDVVELCTASQKSIQFHWKNSNQIKILNGSGFLLTKGPSKLANDRADSRSL 61

QY 87 WDQGNFPLIIKMLKIEDSDTYICEVEDQKEVQLVFGLTANSDTHLQGGSLTLTLESP 146  
 DB 62 WDQGNFPLIIKMLKIEDSDTYICEVEDQKEVQLVFGLTANSDTHLQGGSLTLTLESP 121

QY 147 PSSSSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONQKVEFKIDIVLAFQK 206  
 DB 122 PSSSSVQCRSPRGKNIQGGKTLVSQLELDQSGTWTCTVLONQKVEFKIDIVLAFQK 181

QY 207 ASSIYKKEGEQVEFSFLAFTVEKLTGSGELMWQAEASSSKSWITFDLKNKEYSVKRV 266  
 DB 182 ASSIYKKEGEQVEFSFLAFTVEKLTGSGELMWQAEASSSKSWITFDLKNKEYSVKRV 241

QY 267 TQDPKLQMGKKLPLHLTLPOALPOYAGSGNLTALAEAKTGKLEHVNLVVWRATOLQKNL 326  
 DB 242 TQDPKLQMGKKLPLHLTLPOALPOYAGSGNLTALAEAKTGKLEHVNLVVWRATOLQKNL 301

QY 327 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLNPEAGMWQCLSDSGVLLLESNTKV 386  
 DB 302 TCEVWGPTSPKMLSLKLENKEAKVSKREKAVWVLNPEAGMWQCLSDSGVLLLESNTKV 361

QY 387 LPTWSTPVH 395  
 DB 362 LPTWSTPVH 370

RESULT 94  
 AAR12959  
 ID AAR12959 standard; protein; 370 AA.  
 XX  
 AC AAR12959;  
 XX  
 XX 25-MAR-2003 (revised)  
 DT 09-JAN-2003 (revised)  
 DT 18-SEP-1991 (first entry)  
 XX  
 DE Asn(88), Arg(89), Glu(94) CD4 mutant.  
 XX  
 KM CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;  
 KW human immunodeficiency virus.  
 XX  
 OS Homo sapiens.

```

XX  MO9109123-A.
XX  27-JUN-1991.
XX  15-DEC-1989; 89WO-US005625.
XX  15-DEC-1989; 89WO-US005625.
XX  15-DEC-1989; 89WO-US005625.
XX  (DAND ) DANA FARBER CANCER INST INC.
XX  (REIN/) REINHERZ E L.
XX  Reinherz E, Clayton LK;
XX  WPI; 1991-208149/28.
XX  New soluble human CD4 glyco:protein - has decreased affinity for MHC
XX  class II antigens and is used to treat, prevent and diagnose HIV
XX  infections.
XX  Disclousure; Table 1; 41pp; English.
XX  Mutant M7 is one of 17 modified CD4 proteins analysed for class II MHC
XX  binding. It corresponds to domains I and II of the human CD4 protein in
XX  which Asp(88), Gln(89) and Gln(94) are replaced with Asn, Arg and Glu,
XX  respectively. Mutant M7 is not one of the preferred mutants of the
XX  invention as it retains its Class II MHC binding affinity. See AAR12951-
XX  R12967. (Updated on 09-JAN-2003 to add missing OS field.) (Updated on 25-
XX  MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to correct PA
XX  field.)
XX  Sequence 370 AA;
XX
XX  Query Match 80.8%; Score 1879; DB 2; Length 370;
XX  Best Local Similarity 98.6%; Pred. No. 3.4e-120; Indels 0; Gaps 0;
XX  Matches 364; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
XX
XX  27 KVLGKKGDIVELTCTASQKKSIOFHWKNSNOIKILGNQGSFLTKGPSKLNDRADSRSL 86
XX  2 KVLGKKGDIVELTCTASQKKSIOFHWKNSNOIKILGNQGSFLTKGPSKLNDRADSRSL 61
XX
XX  87 WDOGNFPLIINKLKIEDSDTYICEVEDOKEEVOLVFGLTANSDTHLQGSQSLTLTLESP 146
XX  WDOGNFPLIINKLKIEDSDTYICEVENKKEVELLVFGLTANSDTHLQGSQSLTLTLESP 121
XX  62 WDOGNFPLIINKLKIEDSDTYICEVENKKEVELLVFGLTANSDTHLQGSQSLTLTLESP 121
XX  147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNKVFEFKIDIVLAFOK 206
XX  122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNKVFEFKIDIVLAFOK 181
XX  207 ASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWOAERASSSSKSWITFDLKNKEVSVKRV 266
XX  182 ASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWOAERASSSSKSWITFDLKNKEVSVKRV 241
XX  267 TQDPKLOMGKULPRLHLLTPALPOYAGSGNLTALAKTGKTHQEVNLVVMRATQLOKRL 326
XX  242 TQDPKLOMGKULPRLHLLTPALPOYAGSGNLTALAKTGKTHQEVNLVVMRATQLOKRL 301
XX  327 TCEVWGPTSPKMLSLKLENKEAKVSKREKVPVWVNLNPEAGMOCCLSDSGQVLLLESNIKV 386
XX  302 TCEVWGPTSPKMLSLKLENKEAKVSKREKAVVWVNLNPEAGMOCCLSDSGQVLLLESNIKV 361
XX  387 LPTWSTPVH 395
XX  362 LPTWSTPVH 370
XX
XX  RESULT 95
XX  AAR12962
XX  ID AAR12962 standard; protein; 370 AA.
XX  AC AAR12962;
XX  DT 25-MAR-2003 (revised)

```

```

DT  09-JAN-2003 (revised)
DT  18-SEP-1991 (first entry)
XX
XX  Leu(127), Thr(128), Glu(129) CD4 mutant.
XX  CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;
XX  human immunodeficiency virus.
XX  Homo sapiens.
XX  MO9109123-A.
XX  27-JUN-1991.
XX  15-DEC-1989; 89WO-US005625.
XX  15-DEC-1989; 89WO-US005625.
XX  15-DEC-1989; 89WO-US005625.
XX  (DAND ) DANA FARBER CANCER INST INC.
XX  (REIN/) REINHERZ E L.
XX  Reinherz E, Clayton LK;
XX  WPI; 1991-208149/28.
XX
XX  New soluble human CD4 glyco:protein - has decreased affinity for MHC
XX  class II antigens and is used to treat, prevent and diagnose HIV
XX  infections.
XX  Disclousure; Table 1; 41pp; English.
XX
XX  Mutant M10 is one of 17 modified CD4 proteins analysed for class II MHC
XX  binding. It corresponds to domains I and II of the human CD4 protein in
XX  which Ser(127), Val(128) and Gln(129) are replaced with Leu, Thr and Glu,
XX  respectively. Mutant M10 is not one of the preferred mutants of the
XX  invention as it retains its Class II MHC binding affinity. See AAR12951-
XX  R12967. (Updated on 09-JAN-2003 to add missing OS field.) (Updated on 25-
XX  MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to correct PA
XX  field.)
XX  Sequence 370 AA;
XX
XX  Query Match 80.8%; Score 1879; DB 2; Length 370;
XX  Best Local Similarity 98.6%; Pred. No. 3.4e-120; Indels 0; Gaps 0;
XX  Matches 364; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
XX
XX  27 KVLGKKGDIVELTCTASQKKSIOFHWKNSNOIKILGNQGSFLTKGPSKLNDRADSRSL 86
XX  2 KVLGKKGDIVELTCTASQKKSIOFHWKNSNOIKILGNQGSFLTKGPSKLNDRADSRSL 61
XX
XX  87 WDOGNFPLIINKLKIEDSDTYICEVEDOKEEVOLVFGLTANSDTHLQGSQSLTLTLESP 146
XX  WDOGNFPLIINKLKIEDSDTYICEVEDOKEEVOLVFGLTANSDTHLQGSQSLTLTLESP 121
XX  62 WDOGNFPLIINKLKIEDSDTYICEVEDOKEEVOLVFGLTANSDTHLQGSQSLTLTLESP 121
XX  147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNKVFEFKIDIVLAFOK 206
XX  122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNKVFEFKIDIVLAFOK 181
XX  207 ASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWOAERASSSSKSWITFDLKNKEVSVKRV 266
XX  182 ASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWOAERASSSSKSWITFDLKNKEVSVKRV 241
XX  267 TQDPKLOMGKULPRLHLLTPALPOYAGSGNLTALAKTGKTHQEVNLVVMRATQLOKRL 326
XX  242 TQDPKLOMGKULPRLHLLTPALPOYAGSGNLTALAKTGKTHQEVNLVVMRATQLOKRL 301
XX  327 TCEVWGPTSPKMLSLKLENKEAKVSKREKVPVWVNLNPEAGMOCCLSDSGQVLLLESNIKV 386
XX  302 TCEVWGPTSPKMLSLKLENKEAKVSKREKAVVWVNLNPEAGMOCCLSDSGQVLLLESNIKV 361
XX  387 LPTWSTPVH 395
XX  362 LPTWSTPVH 370
XX
XX

```

```

RESULT 96
ID AAP90992 standard; peptide; 435 AA.
XX
AC AAP90992;
XX
DT 25-MAR-2003 (revised)
DT 04-JUN-1990 (first entry)
XX
DE Human CD4 antigen.
XX
KW Oligopeptide; human CD4 antigen; retrovirus; HIV; therapy; HIV-1; HIV-2;
KW prolylactac; vaccine; diagnostic agent.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 83..89
FT /note="Core sequence"
XX
XX WO8903420-A.
XX
XX 20-APR-1989.
XX
XX 04-AUG-1988; 88WO-US002638.
XX
XX 13-OCT-1987; 87US-00108160.
XX 01-JUN-1988; 88US-00203285.
XX 13-OCT-1988; 88WO-US003592.
XX
XX (GENE-) GENELABS INC.
XX
XX Lifson JD, Hwang K, Elden LE, Nara PL, Fraser B;
XX
XX WPI; 1989-130038/17.
XX
XX Oligopeptide cpds. having partial sequence of human cd4 antigen -
XX inhibiting interaction between CD4 positive cells and retrovirus, esp.
XX HIV.
XX
XX Disclosure; Page 6; 62pp; English.
XX
XX PS The patent concerns oligopeptides derived from the sequence as described
XX CC in the patent title. Compositions of interest will include a core
XX CC sequence (see Features Table), which may be alone, or more usually will
XX CC comprise at least one additional amino acid on at least one end of the
XX CC core sequence. They will contain more than one heteroatom of a bond (not
XX CC a peptide bond) of more than one amino acid joined with a derivatising
XX CC gp. In the preferred sequence, the heteroatom is S and 86-Cys is
XX CC derivatised, and/or the heteroatom is O and 87-Glu is derivatised. The
XX CC derivatising gp. is 7-12C aralkyl opt. contg. 1-2 heteroatoms, esp.
XX CC benzyl or chlorobenzyl. The oligopeptides can be used therapeutically,
XX CC prophylactically or in vitro to modulate response to retroviruses, esp.
XX CC HIV-1 or HIV-2, esp. the fusion of CD4-bearing cells or virion
XX CC infectivity. They are also potentially useful in vaccine development and
XX CC to produce antibodies which are useful as antiviral agents or diagnostic
XX CC reagents. (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-
XX CC MAR-2003 to correct PI field.)
XX
SQ Sequence 435 AA;

```

Query Match 80.8%; Score 1878; DB 1; Length 435;  
Best Local Similarity 87.9%; Pred. No. 4,8e-120;  
Matches 379; Conservative 2; Mismatches 20; Indels 30; Gaps 2;

```

24 GGNKVVLLGKKGPVLTCTASQKKSIOFWKNSNOIKITINGSGFLTKGPSKLNRAISR 83
DB 1 GGNKVVLLGKKGPVLTCTASQKKSIOFWKNSNOIKITINGSGFLTKGPSKLNRAISR 60
84 RSLMGGNFPILIKNLKIEDSDTYICEVEQKEEVQLLVFGLTANSDFLLQGSLTTL 143

```

```

DB 61 RSLTDQGNFPLIKNLKIEDSDTYICEVEQKEEVQLLVFGLTANSDFLLQGSLTTL 120
QY 144 ESPGSSSPVOCRSRGNIGGKTLVSQLELDQSGTWCTTVLONQKVEFKIDIVLA 203
DB 121 ESPGSSSPVOCRSRGNIGGKTLVSQLELDQSGTWCTTVLONQKVEFKIDIVLA 180
QY 204 FOKASSIYKKEGEQVESFPPLAFVTEKLTSGGELMWAERASSSKSWITFDLKNKEYSV 263
DB 181 FOKASSIYKKEGEQVESFPPLAFVTEKLTSGGELMWAERASSSKSWITFDLKNKEYSV 240
QY 264 KRVTQDPFLQWKKLPLHLTLPOALPOYAGSGNLTALAEATGKLHDEVNLVWRAATQLO 323
DB 241 KRVTQDPFLQWKKLPLHLTLPOALPOYAGSGNLTALAEATGKLHDEVNLVWRAATQLO 300
QY 324 KNLTEWVGPTSPKLMLSLKENKEAKYSKREKPYWVUNPFAGMQCLLSGQVLLBSN 383
DB 301 KNLTEWVGPTSPKLMLSLKENKEAKYSKREKPYWVUNPFAGMQCLLSGQVLLBSN 360
QY 384 IKVLPWSTPVHPRASALPAPPTGSALPDQTASALPDPAASALPALAVISFLIGHGL 443
DB 361 IKVLPWSTPVHP-----MALIVGVAGVAGLILFIGL 391
QY 444 GV-ACVLTARTR 453
DB 392 GIFPCVRCRHR 402

```

```

RESULT 97
AAR12966
ID AAR12966 standard; protein; 370 AA.
XX
XX AAR12966;
XX
XX 25-MAR-2003 (revised)
XX 09-JAN-2003 (revised)
XX 18-SEP-1991 (first entry)
XX
XX Thr(162), Leu(163), Asp(164) CD4 mutant.
XX
XX CD4, T lymphocyte; gp120; major histocompatibility complex; AIDS;
XX KW human immunodeficiency virus.
XX
XX OS Homo sapiens.
XX
XX PN WO9109123-A.
XX
XX PD 27-JUN-1991.
XX
XX PF 15-DEC-1989; 89WO-US005625.
XX
XX PR 15-DEC-1989; 89WO-US005625.
XX
XX PA (DANA) FARBER CANCER INST INC.
XX (REIN/) REINHERZ E L.
XX
XX PI Reinherz E, Clayton LK;
XX
XX DR WPI; 1991-208149/28.
XX
XX PT New soluble human CD4 glyco:protein - has decreased affinity for MHC
XX PT class II antigens and is used to treat, prevent and diagnose HIV
XX PT infections.
XX
XX PS Disclosure; Table 1; 41pp; English.
XX
XX CC Mutant M15 is one of 17 modified CD4 proteins analysed for class II MHC
XX CC binding. It corresponds to domains I and II of the human CD4 protein in
XX CC which Leu(162), Glu(163) and Asn(164) are replaced with Thr, Leu and Asp,
XX CC respectively. Mutant M15 is not one of the preferred mutants of the
XX CC invention as it retains its Class II MHC binding affinity. See AAR12951-
XX CC R12967. (Updated on 09-JAN-2003 to add missing OS field.) (Updated on 25-
XX CC MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to correct PA
XX CC field.)

```

```

XX 5Q Sequence 370 AA;
Query Match 80.6%; Score 1874; DB 2; Length 370;
Best Local Similarity 98.6%; Pred. No. 7,4e-120;
Matches 364; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 27 KVLVGGKGDVVELTCTASQKKSIOFHWKNSNQIKILGNQGSFLTGKPSKLNDRADSRSL 86
DB 2 KVLVGGKGDVVELTCTASQKKSIOFHWKNSNQIKILGNQGSFLTGKPSKLNDRADSRSL 61
QY 87 WDGGNFPLIINKLKIEDSDTYICEVEDQKEVQLVFGLTANSSTHLQGSFLTLESP 146
DB 62 WDGGNFPLIINKLKIEDSDTYICEVEDQKEVQLVFGLTANSSTHLQGSFLTLESP 121
QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVQLQNKVEFKIDIVLAFQK 206
DB 122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVQLQNKVEFKIDIVLAFQK 181
QY 207 ASSIVYKKEGEQVEFSPLAFVTEKLTGSGELMMQAEKASSKSWITFDLKNKESVYKRV 266
DB 182 ASSIVYKKEGEQVEFSPLAFVTEKLTGSGELMMQAEKASSKSWITFDLKNKESVYKRV 241
QY 267 TODPKLQMGKKLPLHLTLPOALPOYAGSGNLTALBAKTKLHOEVNLVVMRATQLOKXL 326
DB 242 TODPKLQMGKKLPLHLTLPOALPOYAGSGNLTALBAKTKLHOEVNLVVMRATQLOKXL 301
QY 327 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVVNLNPEAGMOCILSDSGVLLSNIKV 386
DB 302 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVVNLNPEAGMOCILSDSGVLLSNIKV 361
QY 387 LPTWSTPVH 395
DB 362 LPTWSTPVH 370

RESULT 98
AAR12952
ID AAR12952 standard; protein; 370 AA.
XX
AC AAR12952;
XX
DT 25-MAR-2003 (revised)
DT 09-JAN-2003 (revised)
DT 18-SEP-1991 (first entry)
XX
DE Ile(23), Thr(24), Val(25) CD4 mutant.
XX
KW CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;
KW human immunodeficiency virus.
XX
OS Homo sapiens.
XX
PN WO9109123-A.
XX
PD 27-JUN-1991.
XX
PF 15-DEC-1989; 89WO-US005625.
XX
PR 15-DEC-1989; 89WO-US005625.
XX
PA (DAND ) DANA FARBER CANCER INST INC.
PA (REIN/) REINHERZ E L.
XX
PI Reinherz E, Clayton LK;
XX
WPI; 1991-208149/28.
XX
PT New soluble human CD4 glyco:protein - has decreased affinity for MHC
XX infections.
XX
PS Claim 3; Table 1; 41pp; English.

```

```

XX 5Q Sequence 370 AA;
Query Match 80.6%; Score 1873; DB 2; Length 370;
Best Local Similarity 98.6%; Pred. No. 8,7e-120;
Matches 364; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 27 KVLVGGKGDVVELTCTASQKKSIOFHWKNSNQIKILGNQGSFLTGKPSKLNDRADSRSL 86
DB 2 KVLVGGKGDVVELTCTASQKKSIOFHWKNSNQIKILGNQGSFLTGKPSKLNDRADSRSL 61
QY 87 WDGGNFPLIINKLKIEDSDTYICEVEDQKEVQLVFGLTANSSTHLQGSFLTLESP 146
DB 62 WDGGNFPLIINKLKIEDSDTYICEVEDQKEVQLVFGLTANSSTHLQGSFLTLESP 121
QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVQLQNKVEFKIDIVLAFQK 206
DB 122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVQLQNKVEFKIDIVLAFQK 181
QY 207 ASSIVYKKEGEQVEFSPLAFVTEKLTGSGELMMQAEKASSKSWITFDLKNKESVYKRV 266
DB 182 ASSIVYKKEGEQVEFSPLAFVTEKLTGSGELMMQAEKASSKSWITFDLKNKESVYKRV 241
QY 267 TODPKLQMGKKLPLHLTLPOALPOYAGSGNLTALBAKTKLHOEVNLVVMRATQLOKXL 326
DB 242 TODPKLQMGKKLPLHLTLPOALPOYAGSGNLTALBAKTKLHOEVNLVVMRATQLOKXL 301
QY 327 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVVNLNPEAGMOCILSDSGVLLSNIKV 386
DB 302 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVVNLNPEAGMOCILSDSGVLLSNIKV 361
QY 387 LPTWSTPVH 395
DB 362 LPTWSTPVH 370

RESULT 99
AAR12967
ID AAR12967 standard; protein; 370 AA.
XX
AC AAR12967;
XX
DT 25-MAR-2003 (revised)
DT 09-JAN-2003 (revised)
DT 18-SEP-1991 (first entry)
XX
DE Asp(155), Phe(156), Asn(158) CD4 mutant.
XX
KW CD4; T lymphocyte; gp120; major histocompatibility complex; AIDS;
KW human immunodeficiency virus.
XX
OS Homo sapiens.
XX
PN WO9109123-A.
XX
PD 27-JUN-1991.
XX
PF 15-DEC-1989; 89WO-US005625.
XX
PR 15-DEC-1989; 89WO-US005625.
XX
PA (DAND ) DANA FARBER CANCER INST INC.
PA (REIN/) REINHERZ E L.

```

```

XX XX Reinherz E, Clayton LK;
XX XX
XX DR WPI; 1991-208149/28.
XX PT New soluble human CD4 glyco:protein - has decreased affinity for MHC
XX PT class II antigens and is used to treat, prevent and diagnose HIV
XX PT infections.
XX PS Disclosure; Table 1; 41pp; English.
XX XX
XX CC Mutant M14 is one of 17 modified CD4 proteins analysed for class II MHC
XX CC binding. It corresponds to domains I and II of the human CD4 protein in
XX CC which Gly(155), Thr(156) and Thr(158) are replaced with Asp, Phe and Asn,
XX CC respectively. Mutant M14 is not one of the preferred mutants of the
XX CC invention as it could not be evaluated for Class II MHC binding affinity.
XX CC The substitutions grossly affect the structure of the external CD4
XX CC domain so that reactivity with all anti-CD4 monoclonal antibodies tested
XX CC was eliminated. See AAR12951-R12966. (Updated on 09-JAN-2003 to add
XX CC missing OS field.) (Updated on 25-MAR-2003 to correct PR field.) (Updated
XX CC on 25-MAR-2003 to correct PA field.)
XX SQ Sequence 370 AA;

Query Match      80.5%; Score 1872; DB 2; Length 370;
Best Local Similarity 98.6%; Pred. No. 1e-119;
Matches 364; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 27 KVLGKGGDTVELTCTASQKKSIOFHMKNNSNOIKILNGSFLTGKPSKLNDRADSRSL 86
DB 2 KVLGKGGDTVELTCTASQKKSIOFHMKNNSNOIKILNGSFLTGKPSKLNDRADSRSL 61
QY 87 WQGNFPLIIKILKIEDSDTYICEVEDQKEEVQLVFGITANSDFHLQGGSLTTLTSP 146
DB 62 WQGNFPLIIKILKIEDSDTYICEVEDQKEEVQLVFGITANSDFHLQGGSLTTLTSP 121
QY 147 PSSSSVQCRSPRGKNIQGGKTLVSQLELDSDGTCTVLOKQKVEFKIDIVVLAPOK 206
DB 122 PSSSSVQCRSPRGKNIQGGKTLVSQLELDSDGTCTVLOKQKVEFKIDIVVLAPOK 181
QY 207 ASSIYKKEGEQVEFSPLAFTVEKLTGSGELMWOAERASSKSWITFDLKNKEYSVKRV 266
DB 182 ASSIYKKEGEQVEFSPLAFTVEKLTGSGELMWOAERASSKSWITFDLKNKEYSVKRV 241
QY 267 TODPFLQMGKCLPLHLTLPOALPOYAGSGNLTLEAKTGKLGHOEVNLVVWRATQLQ 326
DB 242 TODPFLQMGKCLPLHLTLPOALPOYAGSGNLTLEAKTGKLGHOEVNLVVWRATQLQ 301
QY 327 TCEVWGPTSPKLMLSIKLENKEAKVSKREKPYVWVLPNPEAGMOCCLSDSGOVLLESNTKV 386
DB 302 TCEVWGPTSPKLMLSIKLENKEAKVSKREKPYVWVLPNPEAGMOCCLSDSGOVLLESNTKV 361
QY 387 LFTWSTPVH 395
DB 362 LFTWSTPVH 370

RESULT 100
AAP91289
ID AAP91289 standard; protein; 435 AA.
XX AC
XX AAP91289;
XX DT 25-MAR-2003 (revised)
XX DT 21-JAN-1991 (first entry)
XX DE Sequence of a segment of the human CD4 molecule.
XX KM HIV-1; HIV-2; vaccine; gp 120; diagnosis; retrovirus.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers

```

```

FT Region      83..89
FT FT /note= "core peptide"
XX XX
XX PN W08903813-A.
XX PD 05-MAY-1989.
XX PF 13-OCT-1988; 88WO-US003592.
XX PR 13-OCT-1987; 87US-00108160.
XX PR 01-JUN-1988; 88US-00203285.
XX PA (GENE-) GENELABS INC.
XX PI Lifson J, Hwang K, Eiden LE, Nara PL, Frazer B;
XX DR WPI; 1989-150731/20.
XX PT Anti-retro-viral polypeptide(s) - comprising a sequence of human CD4
XX PT having amino acid joined to deriving gp.
XX PS Disclosure; Table 1, P 7; 67pp; English.
XX XX
XX CC Polypeptide is claimed, comprising a sequence of 7 or more consecutive aa
XX CC of human CD4. Also claimed is a cpd. comprising a sequence of 10 or more
XX CC consecutive aa of human CD4 which includes the core peptide (see FT).
XX CC (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to
XX CC correct PI field.)
XX SQ Sequence 435 AA;

Query Match      80.5%; Score 1871; DB 1; Length 435;
Best Local Similarity 87.7%; Pred. No. 1.4e-119;
Matches 378; Conservative 2; Mismatches 21; Indels 30; Gaps 2;

QY 24 QGNKVLGKGGDTVELTCTASQKKSIOFHMKNNSNOIKILNGSFLTGKPSKLNDRADSR 83
DB 1 QGNKVLGKGGDTVELTCTASQKKSIOFHMKNNSNOIKILNGSFLTGKPSKLNDRADSR 60
QY 84 RSLMDQGNFPLIIKILKIEDSDTYICEVEDQKEEVQLVFGITANSDFHLQGGSLTTL 143
DB 61 RSLMDQGNFPLIIKILKIEDSDTYICEVEDQKEEVQLVFGITANSDFHLQGGSLTTL 120
QY 144 ESPSSSVQCRSPRGKNIQGGKTLVSQLELDSDGTCTVLOKQKVEFKIDIVVLA 203
DB 121 ESPSSSVQCRSPRGKNIQGGKTLVSQLELDSDGTCTVLOKQKVEFKIDIVVLA 180
QY 204 FOKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMWOAERASSKSWITFDLKNKEYSV 263
DB 181 FOKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMWOAERASSKSWITFDLKNKEYSV 240
QY 264 KRVTDPLQMGKCLPLHLTLPOALPOYAGSGNLTLEAKTGKLGHOEVNLVVWRATQLQ 323
DB 241 KRVTDPLQMGKCLPLHLTLPOALPOYAGSGNLTLEAKTGKLGHOEVNLVVWRATQLQ 300
QY 324 KNLTCVWGPTSPKLMLSIKLENKEAKVSKREKPYVWVLPNPEAGMOCCLSDSGOVLLESNTKV 383
DB 301 KNLTCVWGPTSPKLMLSIKLENKEAKVSKREKPYVWVLPNPEAGMOCCLSDSGOVLLESNTKV 360
QY 384 IKVLPTWSTPVHPAPASALPAPPTGSALPDPTASALPDPPASALPALALVISFLGLGL 443
DB 361 IKVLPTWSTPVHP-----MALIVGVAGVAGLLIFGL 391

```

Search completed: August 3, 2004, 13:08:17  
Job time : 52.0872 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 13:14:00 ; Search time 38.3425 Seconds  
(without alignments) 3706.029 Million cell updates/sec

Title: SEQ5  
Perfect score: 2325  
Sequence: 1 MNRGVPRHLLVLVQLALLP.....VISFLGLGLGVACVLAATR 453

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 125 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/ECT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/ECTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2040	87.7	462	10	US-09-939-537-5
2	2040	87.7	462	10	US-09-243-008-5
3	2038.5	87.7	575	10	US-09-939-537-4
4	2038.5	87.7	575	11	US-09-243-008-4
5	2038	87.6	457	11	US-09-891-119A-9
6	2037	87.6	398	10	US-09-939-537-29
7	2037	87.6	532	10	US-09-939-537-6
8	2037	87.6	532	11	US-09-243-008-6
9	2032	87.4	458	14	US-10-103-597A-39
10	2032	87.4	458	14	US-10-188-44A-39
11	2032	87.4	458	14	US-10-207-655-170
12	2024	87.1	458	12	US-10-151-274-3
13	2017	86.8	402	14	US-10-097-044A-1
14	2014	86.6	458	8	US-08-681-219-27
15	2014	86.6	458	11	US-09-230-111C-25

16	2014	86.6	458	14	US-10-092-138-25	Sequence 25, Appl
17	2001	86.1	397	11	US-09-891-119A-2	Sequence 2, Appl
18	1904	81.9	434	14	US-10-097-044A-4	Sequence 4, Appl
19	1899	81.7	370	9	US-09-759-841-6	Sequence 6, Appl
20	1894.5	81.5	448	14	US-10-024-329-32	Sequence 7, Appl
21	1050.5	45.2	310	8	US-08-485-163-7	Sequence 3, Appl
22	1050.5	45.2	310	9	US-0766-995-6	Sequence 6, Appl
23	1046	45.0	530	8	US-08-485-163-5	Sequence 5, Appl
24	1046	45.0	530	9	US-0766-995-4	Sequence 4, Appl
25	1045	44.9	432	8	US-08-485-163-3	Sequence 3, Appl
26	1045	44.9	432	9	US-0766-995-2	Sequence 2, Appl
27	1041	44.8	203	10	US-09-939-537-31	Sequence 31, Appl
28	993	42.7	612	14	US-10-125-692-10	Sequence 10, Appl
29	921	39.6	788	14	US-10-073-118-26	Sequence 26, Appl
30	919	39.5	590	9	US-09-934-060A-13	Sequence 13, Appl
31	919	39.5	720	9	US-09-934-060A-2	Sequence 2, Appl
32	919	39.5	720	9	US-09-934-060A-4	Sequence 4, Appl
33	911	39.2	178	9	US-09-934-060A-26	Sequence 26, Appl
34	881	37.9	184	14	US-10-024-329-33	Sequence 33, Appl
35	494	21.2	94	11	US-09-881-119A-10	Sequence 10, Appl
36	486	20.9	93	11	US-10-105-545-26	Sequence 26, Appl
37	373	16.0	73	9	US-09-929-924-16	Sequence 16, Appl
38	371	16.0	73	9	US-09-929-924-36	Sequence 36, Appl
39	371	16.0	73	9	US-09-929-924-37	Sequence 37, Appl
40	368	15.8	73	9	US-09-929-924-33	Sequence 33, Appl
41	367	15.8	73	9	US-09-929-924-34	Sequence 34, Appl
42	367	15.8	73	9	US-09-929-924-35	Sequence 35, Appl
43	337	14.5	84	9	US-09-135-228B-8	Sequence 8, Appl
44	293	12.6	240	9	US-09-939-165-2	Sequence 2, Appl
45	288	12.4	58	10	US-09-939-537-35	Sequence 35, Appl
46	238	10.2	95	14	US-10-105-545-25	Sequence 25, Appl
47	197	8.5	72	9	US-09-929-924-17	Sequence 17, Appl
48	162	7.0	570	16	US-10-311-823-13	Sequence 13, Appl
49	162	7.0	586	16	US-10-311-823-7	Sequence 7, Appl
50	157	6.8	1315	15	US-10-094-886-52	Sequence 52, Appl
51	157	6.8	1385	15	US-10-094-886-38	Sequence 38, Appl
52	155	6.7	739	13	US-10-047-542-69	Sequence 69, Appl
53	155	6.7	739	14	US-10-207-655-139	Sequence 139, Appl
54	155	6.7	739	14	US-10-234-041-7	Sequence 7, Appl
55	152	6.5	1479	12	US-10-211-956A-325	Sequence 325, Appl
56	152	6.5	1496	12	US-10-211-962-82	Sequence 82, Appl
57	152	6.5	1496	14	US-10-021-660-125	Sequence 125, Appl
58	152	6.5	1496	15	US-10-331-496A-28	Sequence 28, Appl
59	152	6.5	1498	12	US-10-276-774-1957	Sequence 1957, Appl
60	152	6.5	1498	12	US-10-243-552-899	Sequence 899, Appl
61	150	6.5	50	14	US-10-076-674-4	Sequence 4, Appl
62	150	6.5	50	14	US-10-076-674-5	Sequence 5, Appl
63	150	6.5	50	15	US-10-355-161A-5	Sequence 5, Appl
64	150	6.5	50	15	US-10-355-161A-4	Sequence 4, Appl
65	146	6.3	1477	14	US-10-274-583-20	Sequence 20, Appl
66	145	6.2	701	15	US-10-168-417A-4	Sequence 4, Appl
67	143	6.2	647	14	US-10-176-847-32	Sequence 32, Appl
68	143	6.2	647	14	US-10-234-041-6	Sequence 6, Appl
69	143	6.2	5636	14	US-10-032-189-128	Sequence 128, Appl
70	143	6.2	5636	15	US-10-120-801-72	Sequence 72, Appl
71	143	6.2	5636	15	US-10-023-634-93	Sequence 93, Appl
72	143	6.2	5636	16	US-10-408-765A-1895	Sequence 1895, Appl
73	141	6.1	956	15	US-10-108-260A-3037	Sequence 3037, Appl
74	140.5	6.0	737	9	US-09-925-301-1133	Sequence 1133, Appl
75	139.5	6.0	374	16	US-10-311-823-16	Sequence 16, Appl
76	138.5	6.0	442	16	US-10-311-823-12	Sequence 12, Appl
77	138.5	6.0	458	16	US-10-311-823-4	Sequence 4, Appl
78	138	5.9	2828	9	US-09-905-129-21	Sequence 21, Appl
79	138	5.9	2828	9	US-09-991-630-21	Sequence 21, Appl
80	138	5.9	2828	12	US-10-072-012-607	Sequence 607, Appl
81	138	5.9	2828	12	US-10-072-012-608	Sequence 608, Appl
82	138	5.9	2828	12	US-10-454-351-21	Sequence 21, Appl
83	138	5.9	2828	12	US-09-862-318-21	Sequence 21, Appl
84	138	5.9	2828	14	US-10-176-647-54	Sequence 54, Appl
85	138	5.9	2828	14	US-10-177-293-110	Sequence 110, Appl
86	138	5.9	2828	14	US-10-031-822-49	Sequence 49, Appl
87	138	5.9	2828	14	US-10-032-189-126	Sequence 126, Appl
88	138	5.9	2828	15	US-10-295-027-58	Sequence 58, Appl

89	138	5.9	2828	15	US-10-236-027-11175	Sequence 1175, App1
90	137.5	5.9	702	12	US-10-253-286-282	Sequence 282, App1
91	137.5	5.9	702	12	US-10-360-136-16	Sequence 16, App1
92	137.5	5.9	702	14	US-10-157-031-441	Sequence 341, App1
93	137.5	5.9	702	14	US-10-207-655-87	Sequence 87, App1
94	137.5	5.9	702	15	US-10-117-937-592	Sequence 592, App1
95	137.5	5.9	702	15	US-10-295-027-332	Sequence 332, App1
96	137.5	5.9	702	15	US-10-224-871-282	Sequence 282, App1
97	137.5	5.9	734	9	US-09-756-551A-17	Sequence 17, App1
98	137	5.9	5635	16	US-10-451-168-17	Sequence 78, App1
99	136.5	5.9	595	14	US-10-108-698-4871	Sequence 4871, App1
100	136.5	5.9	1395	9	US-09-800-198-56	Sequence 67
101	136.5	5.9	1395	10	US-09-808-776-15	Sequence 56, App1
102	136.5	5.9	1395	14	US-10-280-776-15	Sequence 15, App1
103	136	5.8	512	15	US-10-094-749-4493	Sequence 2493, App1
104	134.5	5.8	1260	13	US-10-024-918-27	Sequence 27, App1
105	134.5	5.8	1260	16	US-10-650-509-27	Sequence 27, App1
106	134.5	5.8	3931	15	US-10-120-801-18	Sequence 18, App1
107	133.5	5.7	6620	15	US-10-080-334-290	Sequence 290, App1
108	133.5	5.7	6620	16	US-10-408-765A-2291	Sequence 2291, App1
109	133.5	5.7	7968	13	US-10-077-130-5	Sequence 5, App1
110	133	5.7	2783	15	US-10-369-493-6344	Sequence 6344, App1
111	132.5	5.7	352	12	US-10-047-865-89	Sequence 89, App1
112	132.5	5.7	352	16	US-10-038-854-156	Sequence 156, App1
113	132.5	5.7	1596	16	US-10-408-765A-992	Sequence 992, App1
114	132.5	5.7	2338	16	US-10-476-397-4	Sequence 4, App1
115	132.5	5.7	4675	15	US-10-093-463-74	Sequence 74, App1
116	132.5	5.7	4691	15	US-10-093-463-74	Sequence 75, App1
117	132	5.7	1011	12	US-09-825-753A-65	Sequence 65, App1
118	131	5.6	1257	15	US-10-116-275-02	Sequence 202, App1
119	131	5.6	1257	15	US-10-411-010-22	Sequence 22, App1
120	131	5.6	1257	15	US-10-411-010-22	Sequence 23, App1
121	131	5.6	5175	15	US-10-120-801-74	Sequence 74, App1
122	131	5.6	5175	15	US-10-369-493-6859	Sequence 6859, App1
123	131	5.6	5185	15	US-10-369-493-6861	Sequence 6861, App1
124	131	5.6	5185	12	US-10-346-863-34	Sequence 34, App1
125	131	5.6	5198	15	US-10-120-801-75	Sequence 75, App1

## ALIGNMENTS

RESULT 1  
 US-09-939-537-5  
 : Sequence 5, Application US/09939537  
 : Publication No. US20030138410A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Seed, Brian  
 : Banapour, Babak  
 : Romeo, Charles  
 : Kolanus, Waldemar  
 : TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED  
 : CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS  
 : NUMBER OF SEQUENCES: 53  
 : CORRESPONDENCE ADDRESSES:  
 : ADDRESSEE: Clark & Elbing LLP  
 : STREET: 176 Federal Street  
 : CITY: Boston  
 : STATE: MA  
 : COUNTRY: USA  
 : ZIP: 02110  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Diskette  
 : COMPUTER: IBM Compatible  
 : OPERATING SYSTEM: DOS  
 : SOFTWARE: PASCSEQ for Windows Version 2.0  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/09/939,537  
 : FILING DATE: 24-Aug-2001  
 : CLASSIFICATION: <Unknown>  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: 08/284,391  
 : FILING DATE: 02-AUG-1994

```

1      APPLICATION NUMBER: 08/195,395
2      FILING DATE: 14-FEB-1994
3      APPLICATION NUMBER: 07/847,566
4      FILING DATE: 06-MAR-1992
5      APPLICATION NUMBER: 07/665,961
6      FILING DATE: 07-MAR-1991
7
8      ATTORNEY/AGENT INFORMATION:
9
10     NAME: Elbing, Karen L
11
12     REGISTRATION NUMBER: 35,238
13     REFERENCE/DOCKET NUMBER: 00786/247001
14
15     TELECOMMUNICATION INFORMATION:
16
17     TELEPHONE: 617-428-0200
18     TELEFAX: 617-428-7045
19
20     TELEX: <Unknown>
21
22     INFORMATION FOR SEQ ID NO: 5:
23
24     SEQUENCE CHARACTERISTICS:
25
26     LENGTH: 462 amino acids
27     TYPE: amino acid
28     STRANDEDNESS: single
29     TOPOLOGY: linear
30
31     MOLECULE TYPE: protein
32
33     SEQUENCE DESCRIPTION: SEQ ID NO: 5:
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
9
```

Query Match	Similarity	97.7%;	Score 2040;	DB 10;	Length 462;
Best Local	Similarity	91.4%;	Pred. No. 1,5e-149;		
Matches	405;	Conservative	1;	Mismatches	9;
				Indels	28;
				Gaps	2
Qy	1	MNRGVPFRHLLVLQLALLPAAIQGNKVILGKGGTVELCTASQKSIQFHWKNSNOIK	60		
Db	1	MNRGVPFRHLLVLQLALLPAAIQGNKVILGKGGTVELCTASQKSIQFHWKNSNOIK	60		
Qy	61	ILGNQGSFLTGPSPKLANDRADSRSLMDQGNFPLIKLKLTIEDSPYICEVEDQKEVQL	120		
Db	61	ILGNQGSFLTGPSPKLANDRADSRSLMDQGNFPLIKLKLTIEDSPYICEVEDQKEVQL	120		
Qy	121	LVFGITANSDFHLIQOGSGLTTLTSPSSPSVOCSPRGKNIQGGKTLTSLVQLDPSG	180		
Db	121	LVFGITANSDFHLIQOGSGLTTLTSPSSPSVOCSPRGKNIQGGKTLTSLVQLDPSG	180		
Qy	181	TWTCITVLQGNQKKEVFEKIDIVLAEQKASSIYYKKEGEQVEFSFPLAFVFEKLTGSGELMW	240		
Db	181	TWTCITVLQGNQKKEVFEKIDIVLAEQKASSIYYKKEGEQVEFSFPLAFVFEKLTGSGELMW	240		
Qy	241	QAEPASSSKSWITFDLKNKEVSVKRVTPDKLQNGKKLPLHLTLPQALPQVAGSNNLTIA	300		
Db	241	QAEPASSSKSWITFDLKNKEVSVKRVTPDKLQNGKKLPLHLTLPQALPQVAGSNNLTIA	300		
Qy	301	LEATRGKLGHOVNLVWNRATOLQNKLTCEWGPSPKMLSLKLENKAKYSKREKPVW	360		
Db	301	LEATRGKLGHOVNLVWNRATOLQNKLTCEWGPSPKMLSLKLENKAKYSKREKPVW	360		
Qy	361	LNPAEGMOCCLSSGOVLLESNKVLPTWSTPYVPYPRASALPAPPTGSAALPPDPTASALP	420		
Db	361	LNPAEGMOCCLSSGOVLLESNKVLPTWSTPYVPYPRASALPAPPTGSAALPPDPTASALP	420		
Qy	421	DPAPASALPALAVISFLGLGL	443		
Db	400	-----LCYILDAIILFLYGIVL	415		

RESULT 2

US-09-243-008-5

; Sequence 5, Application US/09243008

; Publication No. US20040005334A1

GENERAL INFORMATION:

APPLICANT: Seed, Brian et al.

TITLE OF INVENTION: Redirection of Cellular Immunity by Receptor Chimeras

NUMBER OF SEQUENCES: 40

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P. C.

STREET: 225 Franklin Street

CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
 COMPUTER: IBM PS/2 Model 502 or 555X  
 OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)  
 SOFTWARE: Wordperfect (Version 5.0)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/243,008  
 FILING DATE: 02-Feb-1999  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/08/394,176  
 FILING DATE: SEPTEMBER 11, 1995  
 APPLICATION NUMBER: 08/203,866  
 FILING DATE: February 28, 1994  
 APPLICATION NUMBER: 07/847,566  
 FILING DATE: March 6, 1992  
 APPLICATION NUMBER: 07/665,961  
 FILING DATE: March 7, 1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Karen F. Lech, Ph.D  
 REGISTRATION NUMBER: 35,238  
 REFERENCE/DOCKET NUMBER: 00786/270001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 542-5070  
 TELEFAX: (617) 542-8906  
 TELEX: 200154  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 462 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 SEQUENCE DESCRIPTION: SEQ ID NO: 5:  
 US-09-243-008-5

Query Match	87.7%	Score	2040	DB	11	Length	462
Best Local Similarity	91.4%	Pred. No.	1.5e-149				
Matches	405	Conservative	1	Mismatches	9	Indels	26
						Gaps	2
QY	1	MNRGVPFRHLLVLVLQALLPAATQGNKVVLLGKKGDVTELTCTASOKKSIOFMKNSNOIX	60				
Db	1	MNRGVPFRHLLVLVLQALLPAATQGNKVVLLGKKGDVTELTCTASOKKSIOFMKNSNOIX	60				
QY	61	ILNGSGFLTGTGRSKLNDPADRSRLMOGNGRPLIIKXKIKSDPTIYCEVDQKEEYVL	120				
Db	61	ILNGSGFLTGTGRSKLNDPADRSRLMOGNGRPLIIKXKIKSDPTIYCEVDQKEEYVL	120				
QY	121	LVFGLTNSDTHLLOGQSLTTLTLESPPGSSPEVQCRSPRGKNIQGGKTLTSVQLELQDSG	180				
Db	121	LVFGLTNSDTHLLOGQSLTTLTLESPPGSSPEVQCRSPRGKNIQGGKTLTSVQLELQDSG	180				
QY	181	TWTCTVLONOKKVEKIDIVLAFQKASSIVYKKSGEYSEFPFLAFVTEKLTGSGELMW	240				
Db	181	TWTCTVLONOKKVEKIDIVLAFQKASSIVYKKSGEYSEFPFLAFVTEKLTGSGELMW	240				
QY	241	QAEARASSKSWITFPDLKNKEVSVKAVTQBPKLQMGKCLKPLHITLTPQALPQVAGSGLTLTA	300				
Db	241	QAEARASSKSWITFPDLKNKEVSVKAVTQBPKLQMGKCLKPLHITLTPQALPQVAGSGLTLTA	300				
QY	301	LEAKTGKILHGEVNLVVMRATQLOKNTLCEVWGPTSPKMLSLKLENKAVSKREKPVWV	360				
Db	301	LEAKTGKILHGEVNLVVMRATQLOKNTLCEVWGPTSPKMLSLKLENKAVSKREKPVWV	360				
QY	361	LNPEAGMNOCLLSDSGOVLLESNINIKVLPWTSPVHPRASALPAPPTGSAALPPOTASALP	420				
Db	361	LNPEAGMNOCLLSDSGOVLLESNINIKVLPWTSPVHPRASALPAPPTGSAALPPOTASALP	420				
QY	421	DPPAASALPALAVISFLIGLGL	443				
Db	400	-----LCYIIDALIFLGIYL	415				

```

1      RESULT 3
2      US-09-939-537-4
3      ; Sequence 4, Application US/09939537
4      ; Publication No. US20030138410A1
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Seed, Brian
7      ; Banapur, Babak
8      ; Romeo, Charles
9      ; Kolanus, Waldemar
10     ; TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
11     ; CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS
12     ; NUMBER OF SEQUENCES: 53
13     ; CORRESPONDENCE ADDRESS:
14     ; ADDRESSEE: Clark & Elbing LLP
15     ; STREET: 176 Federal Street
16     ; CITY: Boston
17     ; STATE: MA
18     ; COUNTRY: USA
19     ; ZIP: 02110
20     ; COMPUTER READABLE FORM:
21     ; MEDIUM TYPE: Diskette
22     ; COMPUTER: IBM Compatible
23     ; OPERATING SYSTEM: DOS
24     ; SOFTWARE: FastSeq for Windows Version 2.0
25     ; CURRENT APPLICATION DATA:
26     ; APPLICATION NUMBER: US/09/939,537
27     ; FILING DATE: 24-Aug-2001
28     ; CLASSIFICATION: <Unknown>
29     ; PRIOR APPLICATION DATA:
30     ; APPLICATION NUMBER: 08/284,391
31     ; FILING DATE: 02-AUG-1994
32     ; APPLICATION NUMBER: 08/195,395
33     ; FILING DATE: 14-FEB-1994
34     ; APPLICATION NUMBER: 07/847,566
35     ; FILING DATE: 06-MAR-1992
36     ; APPLICATION NUMBER: 07/665,961
37     ; FILING DATE: 07-MAR-1991
38     ; ATTORNEY/AGENT INFORMATION:
39     ; NAME: Elbing, Karen L
40     ; REGISTRATION NUMBER: 35,238
41     ; REFERENCE/DOCKET NUMBER: 00786/247001
42     ; TELECOMMUNICATION INFORMATION:
43     ; TELEPHONE: 617-428-0200
44     ; TELEFAX: 617-428-7045
45     ; TELEX: <Unknown>
46     ; INFORMATION FOR SEQ ID NO: 4:
47     ; SEQUENCE CHARACTERISTICS:
48     ; LENGTH: 575 amino acids
49     ; TYPE: amino acid
50     ; STRANDEDNESS: single
51     ; TOPOLOGY: linear
52     ; MOLECULE TYPE: protein
53     ; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
54     ; US-09-939-537-4

```

Query Match	87.7%	Score 2038.5	DB 10	Length 575
Best Local Similarity	92.1%	Prod. No. 2.6e-149		
Matches 406	Conservative 4	Mismatches 8	Indels 23	Gaps 3
Qy	1	MNRGVPFRHLILVQLALLPAPATGKNVVLGKKGDTELCTASQKKSIOFHMKNSQIK	60	
Db	1	MNRGVPFRHLILVQLALLPAPATGKNVVLGKKGDTELCTASQKKSIOFHMKNSQIK	60	
Qy	61	ILGNQSSFLITGPKSLNDRADSRRLWDQGNFPLIINLKI ESDTYICEVEDQKEEVOL	120	
Db	61	ILGNQSSFLITGPKSLNDRADSRRLWDQGNFPLIINLKI ESDTYICEVEDQKEEVOL	120	
Qy	121	LVFELTANSDPHLLQGOSLTITLESPPGSSPSVQCRRPRGNIOGGTILSVSOLELDSG	180	
Db	121	LVFELTANSDPHLLQGOSLTITLESPPGSSPSVQCRRPRGNIOGGTILSVSOLELDSG	180	

Qy 181 TWCTVLQNKKEFKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGSGELMW 240  
Db 181 TWCTVLQNKKEFKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGSGELMW 240  
Qy 241 QAEKASSSSKSWITFDLKNKEVSVKRVYDQPKLQMGKKLPLHLTLPOALPOYAGSGNLTLLA 300  
Db 241 QAEKASSSSKSWITFDLKNKEVSVKRVYDQPKLQMGKKLPLHLTLPOALPOYAGSGNLTLLA 300  
Qy 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPYVW 360  
Db 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPYVW 360  
Qy 361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTPVH--PR-----ASAL 401  
Db 361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTPVHADPKLCYLLDGLIFVIGYITLALYL 420  
Qy 402 PAPPTGSALPDPOQTASALPDP 422  
Db 421 RAKFSRSA-----ETAANLQDP 437

## RESULT 4

US-09-243-008-4  
Sequence 4, Application US/09243008  
Publication No. US2004000534A1  
GENERAL INFORMATION:  
APPLICANT: Seed, Brian et al.  
TITLE OF INVENTION: Redirection of Cellular Immunity by  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 55SX  
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)  
SOFTWARE: Wordperfect (Version 5.0)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/243.008  
FILING DATE: 02-Feb-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/394.176  
FILING DATE: SEPTEMBER 11, 1995  
APPLICATION NUMBER: 08/203.866  
FILING DATE: February 28, 1994  
APPLICATION NUMBER: 07/847.566  
FILING DATE: March 6, 1992  
APPLICATION NUMBER: 07/665.961  
FILING DATE: March 7, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Karen F. Lech, Ph.D  
REGISTRATION NUMBER: 35,238  
REFERENCE/DOCKET NUMBER: 00786/270001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 575 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-243-008-4  
Query Match 87.7%; Score 2038.5; DB 11; Length 575;  
Best Local Similarity 92.1%; Pred. No. 2.6e-149;

Matches 406; Conservative 4; Mismatches 8; Indels 23; Gaps 3;  
Qy 1 MNGVPPRHLLVLOALALPAATQGNKVYLGKKGDVTELTCTASQKSIQFHMKNNSQIK 60  
Db 1 MNGVPPRHLLVLOALALPAATQGNKVYLGKKGDVTELTCTASQKSIQFHMKNNSQIK 60  
Qy 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120  
Db 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120  
Qy 121 LVFGLTANSDTHLLQGSLLTLTESPPGSSPVQCSPPRKNIOGGKTLISVSQLELDGSG 180  
Db 121 LVFGLTANSDTHLLQGSLLTLTESPPGSSPVQCSPPRKNIOGGKTLISVSQLELDGSG 180  
Qy 181 TWCTVLQNKKEFKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGSGELMW 240  
Db 181 TWCTVLQNKKEFKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGSGELMW 240  
Qy 241 QAEKASSSSKSWITFDLKNKEVSVKRVYDQPKLQMGKKLPLHLTLPOALPOYAGSGNLTLLA 300  
Db 241 QAEKASSSSKSWITFDLKNKEVSVKRVYDQPKLQMGKKLPLHLTLPOALPOYAGSGNLTLLA 300  
Qy 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPYVW 360  
Db 301 LEAKTGKLGHOEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPYVW 360  
Qy 361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTPVH--PR-----ASAL 401  
Db 361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTPVHADPKLCYLLDGLIFVIGYITLALYL 420  
Qy 402 PAPPTGSALPDPOQTASALPDP 422  
Db 421 RAKFSRSA-----ETAANLQDP 437

## RESULT 5

US-09-891-119A-9  
Sequence 9, Application US/09891119A  
Publication No. US20040013683A1  
GENERAL INFORMATION:  
APPLICANT: Maddon, Paul J.  
TITLE OF INVENTION: DERIVATIVES OF SOLUBLE T-4  
FILE REFERENCE: 24577-CY-B  
CURRENT APPLICATION NUMBER: US/09/891.119A  
CURRENT FILING DATE: 2001-06-25  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 9  
LENGTH: 457  
TYPE: PRT  
ORGANISM: human  
US-09-891-119A-9

Query Match 87.7%; Score 2038; DB 11; Length 457;

Best Local Similarity 89.4%; Pred. No. 2.1e-149;  
Matches 406; Conservative 2; Mismatches 16; Indels 30; Gaps 2;

Qy 1 MNGVPPRHLLVLOALALPAATQGNKVYLGKKGDVTELTCTASQKSIQFHMKNNSQIK 60  
Db 1 MNGVPPRHLLVLOALALPAATQGNKVYLGKKGDVTELTCTASQKSIQFHMKNNSQIK 60  
Qy 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120  
Db 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEEVOL 120  
Qy 121 LVFGLTANSDTHLLQGSLLTLTESPPGSSPVQCSPPRKNIOGGKTLISVSQLELDGSG 180  
Db 121 LVFGLTANSDTHLLQGSLLTLTESPPGSSPVQCSPPRKNIOGGKTLISVSQLELDGSG 180  
Qy 181 TWCTVLQNKKEFKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGSGELMW 240  
Db 181 TWCTVLQNKKEFKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGSGELMW 240

Qy	24	QARRASSKSWITFEDLKNKEVSKRATQDPKLOMKKKLPLHLTLPLQALPOYAGSGNLTLA	3000
Db	241	QARRASSKSWITFEDLKNKEVSKRATQDPKLOMKKKLPLHLTLPLQALPOYAGSGNLTLA	3000
Qy	301	LEAKTGKTHQEVNLVVMRATQLOKNTLCEVMGPTSPKMLSLKLENKAAYSKEKEVWV	3600
Db	301	LEAKTGKTHQEVNLVVMRATQLOKNTLCEVMGPTSPKMLSLKLENKAAYSKEKEVWV	3600
Qy	361	LNPDAGMOCCLSDSQVLTLESNIKVLPTWSTFVHPRASALPAPPTGSALPDQTASALP	4200
Db	361	LNPDAGMOCCLSDSQVLTLESNIKVLPTWSTFVHPRASALPAPPTGSALPDQTASALP	3960
Qy	421	DPPAASALPALAVISFLLGLGLGV-ACVLARTR	453
Db	397	-----MALIVLGGVAGLLPLTGLTGATFEVCRCHR	425

```

1      RESULT 6
2      US-09-939-537-29
3      Sequence 29, Application US/09939537
4      Publication No. US20030138410A1
5      GENERAL INFORMATION:
6      APPLICANT: Seed, Brian
7      Banapur, Babak
8      Romeo, Charles
9      Kolanus, Waldemar
10     TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
11     CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
12     NUMBER OF SEQUENCES: 53
13     CORRESPONDENCE ADDRESS:
14     ADDRESSEE: Clark & Ebling LLP
15     STREET: 176 Federal Street
16     CITY: Boston
17     STATE: MA
18     COUNTRY: USA
19     ZIP: 02110
20     COMPUTER READABLE FORM:
21     MEDIUM TYPE: Discrete
22     COMPUTER: IBM Compatible
23     OPERATING SYSTEM: DOS
24     SOFTWARE: Pasteo for Windows Version 2.0
25     CURRENT APPLICATION DATA:
26     APPLICATION NUMBER: US/09/939,537
27     FILING DATE: 24-Aug-2001
28     CLASSIFICATION: <Unknown>
29     PRIOR APPLICATION DATA:
30     APPLICATION NUMBER: 08/284,391
31     FILING DATE: 02-AUG-1994
32     APPLICATION NUMBER: 08/195,395
33     FILING DATE: 14-FEB-1994
34     APPLICATION NUMBER: 07/847,566
35     FILING DATE: 06-MAR-1992
36     APPLICATION NUMBER: 07/665,961
37     FILING DATE: 07-MAR-1991
38     ATTORNEY/AGENT INFORMATION:
39     NAME: Ebling, Karen L
40     REGISTRATION NUMBER: 35,238
41     REFERENCE/DOCKET NUMBER: 00786/247001
42     TELECOMMUNICATION INFORMATION:
43     TELEPHONE: 617-428-0200
44     TELEFAX: 617-428-7045
45     TELEX: <Unknown>
46     INFORMATION FOR SEQ ID NO: 29:
47     SEQUENCE CHARACTERISTICS:
48     LENGTH: 398 amino acids
49     TYPE: amino acid
50     STRANDEDNESS: single
51     TOPOLOGY: linear
52     MOLECULE TYPE: protein
53     SEQUENCE DESCRIPTION: SEQ ID NO: 29:
54
55     US-09-939-537-29
56
57     Query Match      87.6%; Score 2037; DB 10; Length 398;

```

Best Local Similarity	100.0%;	Pred. No. 2,1e-149;			
Matches	395;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0
QY	1	MNRGVPEFHLLLVLTQALLPAPATQGNKVLGKGGDTVEITCTASQKSIQPFMKNSNQIK	60		
Db	1	MNRGVPEFHLLLVLTQALLPAPATQGNKVLGKGGDTVEITCTASQKSIQPFMKNSNQIK	60		
QY	61	ILGNQGSFLTQGPSKLNDRADSRNSLMDQGNPFLIIKVLKLTEDSTPIICEVEDQKEVOL	120		
Db	61	ILGNQGSFLTQGPSKLNDRADSRNSLMDQGNPFLIIKVLKLTEDSTPIICEVEDQKEVOL	120		
QY	121	LVFGITASDPHLLDQGSLLTTLTLEPPSSPSVQCRSRGNKIQGKTLTSVQLDQSG	180		
Db	121	LVFGITASDPHLLDQGSLLTTLTLEPPSSPSVQCRSRGNKIQGKTLTSVQLDQSG	180		
QY	181	TWTCTVLONOKKVEFKIDIVLAFQKASSIYKKEGEVSEFSPLAFVTEKLTGSGELMW	240		
Db	181	TWTCTVLONOKKVEFKIDIVLAFQKASSIYKKEGEVSEFSPLAFVTEKLTGSGELMW	240		
QY	241	QAEARSSSSKSWITPDLKNKEVSVKRVTDQPKLQMGKMLPLHLTLPOALPOVAGSGLTLA	300		
Db	241	QAEARSSSSKSWITPDLKNKEVSVKRVTDQPKLQMGKMLPLHLTLPOALPOVAGSGLTLA	300		
QY	301	LEAKTGKLIHQEVNLTVMKATOLQKNLTCEVWGPTSPKMLSLIKLENKEAKSVKREPVAV	360		
Db	301	LEAKTGKLIHQEVNLTVMKATOLQKNLTCEVWGPTSPKMLSLIKLENKEAKSVKREPVAV	360		
QY	361	LNPEAGMWQCLLSDSGQVLTLESNTKVLPTWGSTPVH	395		
Db	361	LNPEAGMWQCLLSDSGQVLTLESNTKVLPTWGSTPVH	395		

RESULT 7  
 US-09-939-537-6  
 ; Sequence 6, Application US/09939537  
 ; Publication No. US20030138410A1  
 ; GENERAL INFORMATION:  
 APPLICANT: Seed, Brian  
 Banapour, Babak  
 Romeo, Charles  
 Kolanus, Waldemar  
 TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED  
 CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS  
 NUMBER OF SEQUENCES: 53  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Clark & Elbing LLP  
 STREET: 176 Federal Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Discrete  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/939,537  
 FILING DATE: 24-Aug-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/284,391  
 FILING DATE: 02-AUG-1994  
 APPLICATION NUMBER: 08/195,395  
 FILING DATE: 14-FEB-1994  
 APPLICATION NUMBER: 07/847,566  
 FILING DATE: 06-MAR-1992  
 APPLICATION NUMBER: 07/665,961  
 FILING DATE: 07-MAR-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Elbing, Karen L  
 REGISTRATION NUMBER: 35,238  
 REFERENCE/DOCKET NUMBER: 00786/247001

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-428-0200  
TELEFAX: 617-428-7045  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 532 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-09-939-537-6

Query Match 87.6%; Score 2037; DB 10; Length 532;  
Best Local Similarity 100.0%; Pred. No. 3.1e-149;  
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVLTQALLPATQGNKVLGKGDVTELTCTASQKSIQPHMKNNOIK 60  
DB 1 MNRGVPRHLLVLTQALLPATQGNKVLGKGDVTELTCTASQKSIQPHMKNNOIK 60  
QY 61 ILNGSGFLTKGSPKLNDRADSRSLMDQGNPFLIKLKIEDSDTYICEVEDQKEVQL 120  
DB 61 ILNGSGFLTKGSPKLNDRADSRSLMDQGNPFLIKLKIEDSDTYICEVEDQKEVQL 120  
QY 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDG 180  
DB 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDG 180  
QY 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDG 180  
DB 181 TWCTCVLONQKKEFKIDIVLAFOKASSIYKKEGEOVEFSPLAFTVEKLTGSGELMW 240  
QY 181 TWCTCVLONQKKEFKIDIVLAFOKASSIYKKEGEOVEFSPLAFTVEKLTGSGELMW 240  
DB 181 TWCTCVLONQKKEFKIDIVLAFOKASSIYKKEGEOVEFSPLAFTVEKLTGSGELMW 240  
QY 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLPQALPOYAGSGNLTLLA 300  
DB 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLPQALPOYAGSGNLTLLA 300  
QY 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKAVSKREKPVW 360  
DB 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKAVSKREKPVW 360  
QY 361 LNPEAGMOCCLSDSGVLESNIKVLPTWSTPVH 395  
DB 361 LNPEAGMOCCLSDSGVLESNIKVLPTWSTPVH 395

RESULT 8  
US-09-243-008-6  
Sequence 6, Application US/09243008  
Publication No. US2004000534A1  
GENERAL INFORMATION:  
APPLICANT: Seed, Brian et al.  
TITLE OF INVENTION: Redirection of Cellular Immunity by  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESS: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)  
SOFTWARE: Wordperfect (Version 5.0)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/243,008  
FILING DATE: 02-Feb-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/394,176

FILING DATE: SEPTEMBER 11, 1995  
APPLICATION NUMBER: 08/203,866  
FILING DATE: February 28, 1994  
APPLICATION NUMBER: 07/847,566  
FILING DATE: March 6, 1992  
APPLICATION NUMBER: 07/665,961  
FILING DATE: March 7, 1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Karen F. Lech, Ph.D  
REGISTRATION NUMBER: 35,238  
REFERENCE/DOCKET NUMBER: 00786/270001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 532 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-09-243-008-6

Query Match 87.6%; Score 2037; DB 11; Length 532;  
Best Local Similarity 100.0%; Pred. No. 3.1e-149;  
Matches 395; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVLTQALLPATQGNKVLGKGDVTELTCTASQKSIQPHMKNNOIK 60  
DB 1 MNRGVPRHLLVLTQALLPATQGNKVLGKGDVTELTCTASQKSIQPHMKNNOIK 60  
QY 61 ILNGSGFLTKGSPKLNDRADSRSLMDQGNPFLIKLKIEDSDTYICEVEDQKEVQL 120  
DB 61 ILNGSGFLTKGSPKLNDRADSRSLMDQGNPFLIKLKIEDSDTYICEVEDQKEVQL 120  
QY 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDG 180  
DB 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDG 180  
QY 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDG 180  
DB 181 TWCTCVLONQKKEFKIDIVLAFOKASSIYKKEGEOVEFSPLAFTVEKLTGSGELMW 240  
QY 181 TWCTCVLONQKKEFKIDIVLAFOKASSIYKKEGEOVEFSPLAFTVEKLTGSGELMW 240  
DB 181 TWCTCVLONQKKEFKIDIVLAFOKASSIYKKEGEOVEFSPLAFTVEKLTGSGELMW 240  
QY 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLPQALPOYAGSGNLTLLA 300  
DB 241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPHLTLPQALPOYAGSGNLTLLA 300  
QY 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKAVSKREKPVW 360  
DB 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKLENKAVSKREKPVW 360  
QY 361 LNPEAGMOCCLSDSGVLESNIKVLPTWSTPVH 395  
DB 361 LNPEAGMOCCLSDSGVLESNIKVLPTWSTPVH 395

RESULT 9  
US-10-103-597A-39  
Sequence 39, Application US/10103597A  
Publication No. US20030096432A1  
GENERAL INFORMATION:  
APPLICANT: Jakobsen, Bent Karsten  
TITLE OF INVENTION: Screening Methods  
FILE REFERENCE: 102286.142  
CURRENT APPLICATION NUMBER: US/10/103,597A  
CURRENT FILING DATE: 2002-10-17  
PRIOR APPLICATION NUMBER: PCT/GB00/03579  
PRIOR FILING DATE: 2000-09-18  
PRIOR APPLICATION NUMBER: GB 9922352.1  
PRIOR FILING DATE: 1999-09-21  
NUMBER OF SEQ ID NOS: 39  
SOFTWARE: FastSeq for Windows Version 4.0

/ SEQ ID NO 39  
/ LENGTH: 458  
/ TYPE: PRT  
/ ORGANISM: Homo sapiens  
US-10-103-597A-39

Query Match 87.4%; Score 2032; DB 14; Length 458;  
Best Local Similarity 89.2%; Pred. No. 6.1e-149;  
Matches 405; Conservative 2; Mismatches 17; Indels 30; Gaps 2;

```
QY 1 MNRGVPFRHLILVQLALLPAATQGNKVVLGKKGDVELTCTASQKKSIOFHMKNSNQIK 60
DB 1 MNRGVPFRHLILVQLALLPAATQGNKVVLGKKGDVELTCTASQKKSIOFHMKNSNQIK 60
QY 61 ILNGGSPFLTKGSPSKLNDRADSRSLMDQGNPPLIIKNLIKEDSDTYICEVEDQKEEYQL 120
DB 61 ILNGGSPFLTKGSPSKLNDRADSRSLMDQGNPPLIIKNLIKEDSDTYICEVEDQKEEYQL 120
QY 121 LVFGLTANSPTHLLQGGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSPTHLLQGGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCVTLQNOKKVEFKIDIVLAFQKASSIYKKEGEVEFSFPLAFTVEKLTGSGELMW 240
DB 181 TWTCVTLQNOKKVEFKIDIVLAFQKASSIYKKEGEVEFSFPLAFTVEKLTGSGELMW 240
QY 241 QAEBAASSKSWITFDLKNKEVSVKRVTPDPKLQMGKPLHLTLPLPALPOYAGSGNLTIA 300
DB 241 QAEBAASSKSWITFDLKNKEVSVKRVTPDPKLQMGKPLHLTLPLPALPOYAGSGNLTIA 300
QY 301 LEAKTGKLGHEVNLVWMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKYSKREKPYWV 360
DB 301 LEAKTGKLGHEVNLVWMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKYSKREKPYWV 360
QY 361 INPEAGMOCCLSDSGQVLLSESNIKVLPWTGSTPVHPRASALPAPPTGSLPDPQTASALP 420
DB 361 INPEAGMOCCLSDSGQVLLSESNIKVLPWTGSTPVHPRASALPAPPTGSLPDPQTASALP 420
QY 421 DEPPASALPALAVISFLGLGLGV-ACVLTARTR 453
DB 421 DEPPASALPALAVISFLGLGLGV-ACVLTARTR 453
QY 397 -----MALIVYGVAGVGLLFLIGLGFPCVRCRHR 425
DB 397 -----MALIVYGVAGVGLLFLIGLGFPCVRCRHR 425
```

RESULT 10  
US-10-188-444-39  
/ Sequence 39, Application US/10188444  
/ Publication No. US20030104635A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Jakobsen, Bent Karsten  
/ TITLE OF INVENTION: Screening Methods  
/ FILE REFERENCE: 102286.142 (CIP)  
/ CURRENT FILING DATE: 2002-07-02  
/ PRIOR APPLICATION NUMBER: PCT/GB00/03579  
/ PRIOR FILING DATE: 2000-09-18  
/ PRIOR APPLICATION NUMBER: GB 9922352.1  
/ NUMBER OF SEQ ID NOS: 39  
/ SOFTWARE: PaetSEO for Windows Version 4.0  
/ SEQ ID NO 39  
/ LENGTH: 458  
/ TYPE: PRT  
/ ORGANISM: Homo sapiens  
US-10-188-444-39

Query Match 87.4%; Score 2032; DB 14; Length 458;  
Best Local Similarity 89.2%; Pred. No. 6.1e-149;  
Matches 405; Conservative 2; Mismatches 17; Indels 30; Gaps 2;

```
QY 1 MNRGVPFRHLILVQLALLPAATQGNKVVLGKKGDVELTCTASQKKSIOFHMKNSNQIK 60
DB 1 MNRGVPFRHLILVQLALLPAATQGNKVVLGKKGDVELTCTASQKKSIOFHMKNSNQIK 60
```

```
QY 61 ILNGGSPFLTKGSPSKLNDRADSRSLMDQGNPPLIIKNLIKEDSDTYICEVEDQKEEYQL 120
DB 61 ILNGGSPFLTKGSPSKLNDRADSRSLMDQGNPPLIIKNLIKEDSDTYICEVEDQKEEYQL 120
QY 121 LVFGLTANSPTHLLQGGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSPTHLLQGGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCVTLQNOKKVEFKIDIVLAFQKASSIYKKEGEVEFSFPLAFTVEKLTGSGELMW 240
DB 181 TWTCVTLQNOKKVEFKIDIVLAFQKASSIYKKEGEVEFSFPLAFTVEKLTGSGELMW 240
QY 241 QAEBAASSKSWITFDLKNKEVSVKRVTPDPKLQMGKPLHLTLPLPALPOYAGSGNLTIA 300
DB 241 QAEBAASSKSWITFDLKNKEVSVKRVTPDPKLQMGKPLHLTLPLPALPOYAGSGNLTIA 300
QY 301 LEAKTGKLGHEVNLVWMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKYSKREKPYWV 360
DB 301 LEAKTGKLGHEVNLVWMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKYSKREKPYWV 360
QY 361 INPEAGMOCCLSDSGQVLLSESNIKVLPWTGSTPVHPRASALPAPPTGSLPDPQTASALP 420
DB 361 INPEAGMOCCLSDSGQVLLSESNIKVLPWTGSTPVHPRASALPAPPTGSLPDPQTASALP 420
QY 421 DEPPASALPALAVISFLGLGLGV-ACVLTARTR 453
DB 421 DEPPASALPALAVISFLGLGLGV-ACVLTARTR 453
QY 397 -----MALIVYGVAGVGLLFLIGLGFPCVRCRHR 425
DB 397 -----MALIVYGVAGVGLLFLIGLGFPCVRCRHR 425
```

RESULT 11  
US-10-207-655-170  
/ Sequence 170, Application US/10207655  
/ Publication No. US20030118592A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Ledbetter, Jeffrey A.  
/ TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS  
/ FILE REFERENCE: 39069.401C1  
/ CURRENT FILING DATE: 2002-07-25  
/ NUMBER OF SEQ ID NOS: 426  
/ SOFTWARE: PatentIn version 3.0  
/ SEQ ID NO 170  
/ LENGTH: 458  
/ TYPE: PRT  
/ ORGANISM: Homo sapiens  
US-10-207-655-170

Query Match 87.4%; Score 2032; DB 14; Length 458;  
Best Local Similarity 89.2%; Pred. No. 6.1e-149;  
Matches 405; Conservative 2; Mismatches 17; Indels 30; Gaps 2;

```
QY 1 MNRGVPFRHLILVQLALLPAATQGNKVVLGKKGDVELTCTASQKKSIOFHMKNSNQIK 60
DB 1 MNRGVPFRHLILVQLALLPAATQGNKVVLGKKGDVELTCTASQKKSIOFHMKNSNQIK 60
QY 61 ILNGGSPFLTKGSPSKLNDRADSRSLMDQGNPPLIIKNLIKEDSDTYICEVEDQKEEYQL 120
DB 61 ILNGGSPFLTKGSPSKLNDRADSRSLMDQGNPPLIIKNLIKEDSDTYICEVEDQKEEYQL 120
QY 121 LVFGLTANSPTHLLQGGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 121 LVFGLTANSPTHLLQGGSLTTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY 181 TWTCVTLQNOKKVEFKIDIVLAFQKASSIYKKEGEVEFSFPLAFTVEKLTGSGELMW 240
DB 181 TWTCVTLQNOKKVEFKIDIVLAFQKASSIYKKEGEVEFSFPLAFTVEKLTGSGELMW 240
QY 241 QAEBAASSKSWITFDLKNKEVSVKRVTPDPKLQMGKPLHLTLPLPALPOYAGSGNLTIA 300
DB 241 QAEBAASSKSWITFDLKNKEVSVKRVTPDPKLQMGKPLHLTLPLPALPOYAGSGNLTIA 300
QY 301 LEAKTGKLGHEVNLVWMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKYSKREKPYWV 360
DB 301 LEAKTGKLGHEVNLVWMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKYSKREKPYWV 360
```

```

Db      301 LEATGKLGHOENVLVVMMRATOLQKLTCEVWGPTSPKMLSLKLENKAKVSKKEKAVMV 360
Qy      361 LNPERGMMQCLSDSGVLTLSNITKVLPTWSTPHPRASALPAPPTSGALDPQTASALP 420
Db      361 LNPERGMMQCLSDSGVLTLSNITKVLPTWSTPHPRASALPAPPTSGALDPQTASALP 396
Qy      421 DDPASALPALAVISFLGLGLGV-ACVLARTR 453
Db      397 -----MALIVGAGVAGLLFLIGLGIFFVCRHR 425

```

## RESULT 12

```

US-10-151-274-3
; Sequence 3, Application US/10151274
; Publication No. US20030064071A1
; GENERAL INFORMATION:
; APPLICANT: Litchman, Dan R.
; APPLICANT: Kwon, Douglas S.
; APPLICANT: van Kooyk, Yvette
; APPLICANT: Geljtenbeck, Theo
; TITLE OF INVENTION: METHODS OF USING A FACILITATOR OF RETROVIRAL ENTRY
; TITLE OF INVENTION: INTO
; FILE REFERENCE: 1049-1-017
; CURRENT APPLICATION NUMBER: US/10/151,274
; PRIOR FILING DATE: 2002-05-20
; PRIOR APPLICATION NUMBER: US/09/517,605
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-151-274-3

```

Query Match 87.1%; Score 2024; DB 12; Length 458;  
 Best Local Similarity 89.0%; Pred. No. 2,6e-148;  
 Matches 404; Conservative 2; Mismatches 18; Indels 30; Gaps 2;

```

Qy      1 MNRGVPFRHLLVQLALPAAATQGNKRVLGKGDVVELTCTASOKKSIOFHMKNNSQIK 60
Db      1 MNRGVPFRHLLVQLALPAAATQGNKRVLGKGDVVELTCTASOKKSIOFHMKNNSQIK 60
Qy      61 ILNGSGFLTKGSPSKLNDRAISRSLMDQGNFPLITIKLKIEDSDTYICEVEDQKEEVOL 120
Db      61 ILNGSGFLTKGSPSKLNDRAISRSLMDQGNFPLITIKLKIEDSDTYICEVEDQKEEVOL 120
Qy      121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 180
Db      121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 180
Qy      181 TWTCVULONOKKVEFKIDIVVLAFOKASSIVYKKEGEVEFSPLATFVEKLTSSGELMW 240
Db      181 TWTCVULONOKKVEFKIDIVVLAFOKASSIVYKKEGEVEFSPLATFVEKLTSSGELMW 240
Qy      241 QAERASSSSKSWITFDLKKKEVSVKRVTDPKLQMGSKPLHLTLPOALPOYAGSGNLTIA 300
Db      241 QAERASSSSKSWITFDLKKKEVSVKRVTDPKLQMGSKPLHLTLPOALPOYAGSGNLTIA 300
Qy      301 LEAKTGKLGHOENVLVVMMRATOLQKLTCEVWGPTSPKMLSLKLENKAKVSKKEKAVMV 360
Db      301 LEAKTGKLGHOENVLVVMMRATOLQKLTCEVWGPTSPKMLSLKLENKAKVSKKEKAVMV 360
Qy      361 LNPERGMMQCLSDSGVLTLSNITKVLPTWSTPHPRASALPAPPTSGALDPQTASALP 420
Db      361 LNPERGMMQCLSDSGVLTLSNITKVLPTWSTPHPRASALPAPPTSGALDPQTASALP 396
Qy      421 DDPASALPALAVISFLGLGLGV-ACVLARTR 453
Db      397 -----MALIVGAGVAGLLFLIGLGIFFVCRHR 425

```

## RESULT 13

US-10-097-044A-1

; Sequence 1, Application US/10097044A  
 ; Publication No. US20030143220A1

## GENERAL INFORMATION:

APPLICANT: Capon, Daniel J.

Gregory, Timothy J.

TITLE OF INVENTION: Adhesion Variants

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSER: Genentech, Inc.

STREET: 460 Point San Bruno Blvd

CITY: South San Francisco

STATE: California

COUNTRY: USA

ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: patin (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/097,044A

FILING DATE: 28-May-2002

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/457,918

FILING DATE: 1-JUN-1995

APPLICATION NUMBER: 08/236311

FILING DATE: 02-MAY-1994

APPLICATION NUMBER: 07/936190

FILING DATE: 26-AUG-1992

APPLICATION NUMBER: 07/842777

FILING DATE: 18-FEB-1992

APPLICATION NUMBER: 07/250785

FILING DATE: 28-SEP-1988

APPLICATION NUMBER: 07/104329

FILING DATE: 02-OCT-1987

ATTORNEY/AGENT INFORMATION:

NAME: Kubinec, Jeffrey S.

REGISTRATION NUMBER: 35,575

REFERENCE/DOCKET NUMBER: P0444P1C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-8228

TELEFAX: 415/952-9881

TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 402 amino acids

TYPE: amino acid

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-10-097-044A-1

Query Match 86.8%; Score 2017; DB 14; Length 402;

Best Local Similarity 99.7%; Pred. No. 7,4e-148;

Matches 392; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy      1 MNRGVPFRHLLVQLALPAAATQGNKRVLGKGDVVELTCTASOKKSIOFHMKNNSQIK 60
Db      1 MNRGVPFRHLLVQLALPAAATQGNKRVLGKGDVVELTCTASOKKSIOFHMKNNSQIK 60
Qy      61 ILNGSGFLTKGSPSKLNDRAISRSLMDQGNFPLITIKLKIEDSDTYICEVEDQKEEVOL 120
Db      61 ILNGSGFLTKGSPSKLNDRAISRSLMDQGNFPLITIKLKIEDSDTYICEVEDQKEEVOL 120
Qy      121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 180
Db      121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSG 180
Qy      181 TWTCVULONOKKVEFKIDIVVLAFOKASSIVYKKEGEVEFSPLATFVEKLTSSGELMW 240
Db      181 TWTCVULONOKKVEFKIDIVVLAFOKASSIVYKKEGEVEFSPLATFVEKLTSSGELMW 240

```



Db 181 TWTCVLQNKQKVEFKIDIVLAFQKASSIYKKEGEVEFSPLAFTVEKLTGSGELMW 240  
QY 241 QAERASSSSKSWITFDLKNKEVSVKRVTPDPLQNGKKLPLHLTLPOALPOYAGSGNLTIA 300  
Db 241 QAERASSSSKSWITFDLKNKEVSVKRVTPDPLQNGKKLPLHLTLPOALPOYAGSGNLTIA 300  
QY 301 LEAKTGKLGHOENVLVWMRAVATQLOKRLTCEVWGPTSPKMLSLKLENKAKYSKREKAVW 360  
Db 301 LEAKTGKLGHOENVLVWMRAVATQLOKRLTCEVWGPTSPKMLSLKLENKAKYSKREKAVW 360  
QY 361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTP 393  
Db 361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTP 393

RESULT 14  
US-08-681-219-27  
; Sequence 27, Application US/08681219  
; Publication No. US20020058607A1  
; GENERAL INFORMATION:  
; APPLICANT: Takeaki Sato and Junn Yanagisawa  
; TITLE OF INVENTION: COMPOUNDS THAT INHIBIT THE INTERACTION BETWEEN  
; TITLE OF INVENTION: SIGNAL-TRANSDUCING PROTEINS AND THE GIGF  
; TITLE OF INVENTION: (PDZ/DHR) DOMAIN AND USES THEREOF  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/681,219  
; FILING DATE: 22-JUL-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 0575/48962/JPM/JKM  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0525  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 458 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-681-219-27

Query Match 86.6%; Score 2014; DB 8; Length 458;  
Best Local Similarity 88.5%; Pred. No. 1.5e-147;  
Matches 402; Conservative 3; Mismatches 19; Indels 30; Gaps 2;

QY 1 MNRGVPPRHLLLVQLALLPATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNNSQIK 60  
Db 1 MNRGVPPRHLLLVQLALLPATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNNSQIK 60  
QY 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNFPLIINKUKIEDSDTYICEVEDQKEEYQL 120  
Db 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNFPLIINKUKIEDSDTYICEVEDQKEEYQL 120  
QY 121 LVFGLTANSDBTHLQGGSLTTLTSSPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
Db 121 LVFGLTANSDBTHLQGGSLTTLTSSPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180

QY 181 TWTCVLQNKQKVEFKIDIVLAFQKASSIYKKEGEVEFSPLAFTVEKLTGSGELMW 240  
Db 181 TWTCVLQNKQKVEFKIDIVLAFQKASSIYKKEGEVEFSPLAFTVEKLTGSGELMW 240  
QY 241 QAERASSSSKSWITFDLKNKEVSVKRVTPDPLQNGKKLPLHLTLPOALPOYAGSGNLTIA 300  
Db 241 QAERASSSSKSWITFDLKNKEVSVKRVTPDPLQNGKKLPLHLTLPOALPOYAGSGNLTIA 300  
QY 301 LEAKTGKLGHOENVLVWMRAVATQLOKRLTCEVWGPTSPKMLSLKLENKAKYSKREKAVW 360  
Db 301 LEAKTGKLGHOENVLVWMRAVATQLOKRLTCEVWGPTSPKMLSLKLENKAKYSKREKAVW 360  
QY 361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTPYPHPRASALPAPPTGSALPDQTASALP 420  
Db 361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTPYPHPRASALPAPPTGSALPDQTASALP 420  
QY 421 DPAPASALPALAVISFLIGIGV-ACVLARTR 453  
Db 397 -----MALIVGVAAGLLIFIGIGIFCVRCHR 425

RESULT 15  
US-09-230-111C-25  
; Sequence 25, Application US/09230111C  
; Publication No. US20030203414A1  
; GENERAL INFORMATION:  
; APPLICANT: Sato, Taka-Aki  
; APPLICANT: Yanagisawa, Junn  
; TITLE OF INVENTION: COMPOUNDS THAT INHIBIT INTERACTION BETWEEN  
; TITLE OF INVENTION: SIGNAL-TRANSDUCING PROTEINS AND THE GIGF (PDZ/DHR)  
; TITLE OF INVENTION: DOMAIN AND USES THEREOF  
; FILE REFERENCE: 48962-A-PCT-US  
; CURRENT APPLICATION NUMBER: US/09/230,111C  
; CURRENT FILING DATE: 1999-05-17  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 25  
; LENGTH: 458  
; TYPE: PRT  
; ORGANISM: human  
US-09-230-111C-25

Query Match 86.6%; Score 2014; DB 11; Length 458;  
Best Local Similarity 88.5%; Pred. No. 1.5e-147;  
Matches 402; Conservative 3; Mismatches 19; Indels 30; Gaps 2;

QY 1 MNRGVPPRHLLLVQLALLPATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNNSQIK 60  
Db 1 MNRGVPPRHLLLVQLALLPATQGNKVVLGKGGDTVELTCTASQKKSIOFHMKNNSQIK 60  
QY 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNFPLIINKUKIEDSDTYICEVEDQKEEYQL 120  
Db 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNFPLIINKUKIEDSDTYICEVEDQKEEYQL 120  
QY 121 LVFGLTANSDBTHLQGGSLTTLTSSPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
Db 121 LVFGLTANSDBTHLQGGSLTTLTSSPPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
QY 181 TWTCVLQNKQKVEFKIDIVLAFQKASSIYKKEGEVEFSPLAFTVEKLTGSGELMW 240  
Db 181 TWTCVLQNKQKVEFKIDIVLAFQKASSIYKKEGEVEFSPLAFTVEKLTGSGELMW 240  
QY 241 QAERASSSSKSWITFDLKNKEVSVKRVTPDPLQNGKKLPLHLTLPOALPOYAGSGNLTIA 300  
Db 241 QAERASSSSKSWITFDLKNKEVSVKRVTPDPLQNGKKLPLHLTLPOALPOYAGSGNLTIA 300  
QY 301 LEAKTGKLGHOENVLVWMRAVATQLOKRLTCEVWGPTSPKMLSLKLENKAKYSKREKAVW 360  
Db 301 LEAKTGKLGHOENVLVWMRAVATQLOKRLTCEVWGPTSPKMLSLKLENKAKYSKREKAVW 360  
QY 361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTPYPHPRASALPAPPTGSALPDQTASALP 420  
Db 361 LNPEAGMOCCLSDSGQVLESNIKVLPTWSTPYPHPRASALPAPPTGSALPDQTASALP 420

Qy 421 DPASALPALAVISFLGLGV-ACVLART 453  
 Db 397 -----MALIVGVAGLLFLGLGIFFCVRCRHR 425

## RESULT 16

US-10-092-138-25  
 ; Sequence 25, Application US/10092138  
 ; Publication No. US20030170723A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sato, Taka-Aki  
 ; TITLE OF INVENTION: METHOD OF PREPARING A PROTEIN ARRAY BASED ON  
 ; FILE REFERENCE: 65823/JPN/PT  
 ; CURRENT APPLICATION NUMBER: US/10/092.138  
 ; CURRENT FILING DATE: 2002-09-06  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 25  
 ; LENGTH: 458  
 ; TYPE: PRP  
 ; ORGANISM: human  
 US-10-092-138-25

Query Match 86.6%; Score 2014; DB 14; Length 458;  
 Best Local Similarity 88.5%; Pred. No. 1,56-147;  
 Matches 402; Conservative 3; Mismatches 19; Indels 30; Gaps 2;

Qy 1 MNRGVPFRHLLVLQALPPAATGKNVYLGKGDYVELTCTASQKSIQPHMKNNOIK 60  
 Db 1 MNRGVPFRHLLVLQALPPAATGKNVYLGKGDYVELTCTASQKSIQPHMKNNOIK 60  
 Qy 61 ILGNQGSFLTKGSKINDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVOL 120  
 Db 61 ILGNQGSFLTKGSKINDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVOL 120  
 Qy 121 LVFGLTANSPTHLQOGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDG 180  
 Db 121 LVFGLTANSPTHLQOGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDG 180  
 Qy 181 TWCTVLOQKKEFKIDIVLAFQKASSIVYKKEGEVESPPLAFTVEKLTGSGELMW 240  
 Db 181 TWCTVLOQKKEFKIDIVLAFQKASSIVYKKEGEVESPPLAFTVEKLTGSGELMW 240  
 Qy 241 QAERASSKSWITFDLKNKEVSVKRVTPDKLQMGKPLHLTLPPALPOYAGSGNLTLA 300  
 Db 241 QAERASSKSWITFDLKNKEVSVKRVTPDKLQMGKPLHLTLPPALPOYAGSGNLTLA 300  
 Qy 301 LEATGKLGHOENVLVVMRATOLQKNTLCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 360  
 Db 301 LEATGKLGHOENVLVVMRATOLQKNTLCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 360  
 Qy 361 LNPEAGMWQCLLSDSGVLLSESNIKVLPWSTPVP 420  
 Db 361 LNPEAGMWQCLLSDSGVLLSESNIKVLPWSTPVP 420  
 Qy 421 DPASALPALAVISFLGLGV-ACVLART 453  
 Db 397 -----MALIVGVAGLLFLGLGIFFCVRCRHR 425

## RESULT 17

US-09-891-119A-2  
 ; Sequence 2, Application US/09891119A  
 ; Publication No. US20040013683A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Magdon, Paul J.  
 ; TITLE OF INVENTION: DERIVATIVES OF SOLUBLE T-4  
 ; FILE REFERENCE: 24577-CY-B  
 ; CURRENT APPLICATION NUMBER: US/09/891.119A  
 ; CURRENT FILING DATE: 2001-06-25  
 ; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 2  
 ; LENGTH: 397  
 ; TYPE: PRP  
 ; ORGANISM: Human  
 US-09-891-119A-2

Query Match 86.1%; Score 2001; DB 11; Length 397;  
 Best Local Similarity 98.7%; Pred. No. 1,36-146;  
 Matches 389; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MNRGVPFRHLLVLQALPPAATGKNVYLGKGDYVELTCTASQKSIQPHMKNNOIK 60  
 Db 1 MNRGVPFRHLLVLQALPPAATGKNVYLGKGDYVELTCTASQKSIQPHMKNNOIK 60  
 Qy 61 ILGNQGSFLTKGSKINDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVOL 120  
 Db 61 ILGNQGSFLTKGSKINDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVOL 120  
 Qy 121 LVFGLTANSPTHLQOGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDG 180  
 Db 121 LVFGLTANSPTHLQOGSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDG 180  
 Qy 181 TWCTVLOQKKEFKIDIVLAFQKASSIVYKKEGEVESPPLAFTVEKLTGSGELMW 240  
 Db 181 TWCTVLOQKKEFKIDIVLAFQKASSIVYKKEGEVESPPLAFTVEKLTGSGELMW 240  
 Qy 241 QAERASSKSWITFDLKNKEVSVKRVTPDKLQMGKPLHLTLPPALPOYAGSGNLTLA 300  
 Db 241 QAERASSKSWITFDLKNKEVSVKRVTPDKLQMGKPLHLTLPPALPOYAGSGNLTLA 300  
 Qy 301 LEATGKLGHOENVLVVMRATOLQKNTLCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 360  
 Db 301 LEATGKLGHOENVLVVMRATOLQKNTLCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 360  
 Qy 361 LNPEAGMWQCLLSDSGVLLSESNIKVLPWSTPVP 394  
 Db 361 LNPEAGMWQCLLSDSGVLLSESNIKVLPWSTPVP 394

## RESULT 18

US-10-097-044A-4  
 ; Sequence 4, Application US/10097044A  
 ; Publication No. US20030143220A1  
 ; GENERAL INFORMATION:

APPLICANT: Capon, Daniel J.  
 Gregory, Timothy J.  
 TITLE OF INVENTION: Adhesion Variants  
 NUMBER OF SEQUENCES: 25  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Genentech, Inc.  
 STREET: 460 Point San Bruno Blvd  
 CITY: South San Francisco  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94080

## COMPUTER READABLE FORM:

MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: patin (Genentech)

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/097.044A  
 FILING DATE: 28-May-2002  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/457.918  
 FILING DATE: 1-JUN-1995  
 APPLICATION NUMBER: 08/236311  
 FILING DATE: 02-MAY-1994  
 APPLICATION NUMBER: 07/936190  
 FILING DATE: 26-AUG-1992  
 APPLICATION NUMBER: 07/842777

```

; FILING DATE: 18-FEB-1992
; APPLICATION NUMBER: 07/250795
; FILING DATE: 28-SEP-1988
; APPLICATION NUMBER: 07/104329
; FILING DATE: 02-OCT-1987
; ATTORNEY/AGENT INFORMATION:
;   NAME: Kubinec, Jeffrey S.
;   REGISTRATION NUMBER: 36,575
;   REFERENCE/DOCKET NUMBER: P0444P1C3
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415/225-8228
;   TELEFAX: 415/952-9881
;   TELE: 910/371-7168
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 434 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-097-044A-4

```

```

Query Match      81.9%; Score 1904; DB 14; Length 434;
Best Local Similarity 99.7%; Pred. No. 4,8e-139;
Matches 369; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 24 QGNKVLGKGGDTVELTCTASQKSIQFHWNKSNQIKILNGSFLLTKGPSKLNDRADSR 83
DB 56 QGNKVLGKGGDTVELTCTASQKSIQFHWNKSNQIKILNGSFLLTKGPSKLNDRADSR 115
QY 84 RSLMDQGNFPLIINLKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLQGGSLTTLT 143
DB 116 RSLMDQGNFPLIINLKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLQGGSLTTLT 175
QY 144 EBPSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNKVEFKIDIVLA 203
DB 176 EBPSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNKVEFKIDIVLA 235
QY 204 FOKASSIYKKEGQVEFSFPLAFTVEKLTSGGELMWAERASSSSKSWITPDLKNKEVS 263
DB 236 FOKASSIYKKEGQVEFSFPLAFTVEKLTSGGELMWAERASSSSKSWITPDLKNKEVS 295
QY 264 KRVTDPLQMGKKLPHLTLPOALPOYAGSGNLTLEAKTGKLEHVNLMVWRATOLQ 323
DB 296 KRVTDPLQMGKKLPHLTLPOALPOYAGSGNLTLEAKTGKLEHVNLMVWRATOLQ 355
QY 324 KNLTCGVWGPTSPKLMLSIKLENKAKVSKREKPVWLNPEAGMOCCLSDSGVLLSEN 383
DB 356 KNLTCGVWGPTSPKLMLSIKLENKAKVSKREKAVWLNPEAGMOCCLSDSGVLLSEN 415
QY 384 IKVLPTWSTP 393
DB 416 IKVLPTWSTP 425

```

```

RESULT 19
US-09-759-841-6
; Sequence 6, Application US/09759841
; Patent No. US20010039026A1
; GENERAL INFORMATION:
;   APPLICANT: Rickett, Graham A
;   APPLICANT: Dobbs, Susan
;   APPLICANT: Perrot, Marcuseos
;   TITLE OF INVENTION: Assay Method
;   FILE REFERENCE: PC10348APME
;   CURRENT APPLICATION NUMBER: US/09/759,841
;   PRIOR FILING DATE: 2001-01-12
;   PRIOR APPLICATION NUMBER: GB 0000661.9
;   PRIOR FILING DATE: 2000-01-12
;   PRIOR APPLICATION NUMBER: GB 0000663.5
;   PRIOR FILING DATE: 2000-01-12
;   PRIOR APPLICATION NUMBER: GB 0000659.3
;   PRIOR FILING DATE: 2000-01-12
;   NUMBER OF SEQ ID NOS: 6

```

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-841-6

```

```

Query Match      81.7%; Score 1899; DB 9; Length 370;
Best Local Similarity 99.7%; Pred. No. 9.3e-139;
Matches 368; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 27 KVLGKGGDTVELTCTASQKSIQFHWNKSNQIKILNGSFLLTKGPSKLNDRADSRSL 86
DB 2 KVLGKGGDTVELTCTASQKSIQFHWNKSNQIKILNGSFLLTKGPSKLNDRADSRSL 61
QY 87 WDQGNFPLIINLKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLQGGSLTTLTSP 146
DB 62 WDQGNFPLIINLKIEDSDTYICEVEDQKEEVQLVFGLTANSPTHLQGGSLTTLTSP 121
QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNKVEFKIDIVLAFOK 206
DB 122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWTCTVLQNKVEFKIDIVLAFOK 181
QY 207 ASSIYKKEGQVEFSFPLAFTVEKLTSGGELMWAERASSSSKSWITPDLKNKEVSARV 266
DB 182 ASSIYKKEGQVEFSFPLAFTVEKLTSGGELMWAERASSSSKSWITPDLKNKEVSARV 241
QY 267 TQDPLQMGKKLPHLTLPOALPOYAGSGNLTLEAKTGKLEHVNLMVWRATOLQK 326
DB 242 TQDPLQMGKKLPHLTLPOALPOYAGSGNLTLEAKTGKLEHVNLMVWRATOLQK 301
QY 327 TCEVWGPTSPKLMLSIKLENKAKVSKREKPVWLNPEAGMOCCLSDSGVLLSENITK 386
DB 302 TCEVWGPTSPKLMLSIKLENKAKVSKREKAVWLNPEAGMOCCLSDSGVLLSENITK 361
QY 387 LPTWSTPVH 395
DB 362 LPTWSTPVH 370

```

```

RESULT 20
US-10-024-329-32
; Sequence 32, Application US/10024329
; Publication No. US20030157063A1
; GENERAL INFORMATION:
;   APPLICANT: SANHADJI, Kamel
;   APPLICANT: TOURAINE, Jean-Louis
;   APPLICANT: LEROY, Pierre
;   APPLICANT: MEHTALI, Majid
;   TITLE OF INVENTION: Gene therapy using anti-gp41 antibody and cd4 immunoadhesin
;   FILE REFERENCE: 109993
;   CURRENT APPLICATION NUMBER: US/10/024,329
;   CURRENT FILING DATE: 2001-12-21
;   NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 448
; TYPE: PRT
; ORGANISM: human bcd4
US-10-024-329-32

```

```

Query Match      81.5%; Score 1894.5; DB 14; Length 448;
Best Local Similarity 87.0%; Pred. No. 2.7e-138;
Matches 395; Conservative 2; Mismatches 18; Indels 39; Gaps 10;

```

```

QY 1 MNRGVFPHLLVYQALPLPATQGNKVLGKGGDTVELTCTASQKSIQFHWNKSNQIK 60
DB 1 MNRGVFPHLLVYQALPLPATQGNKVLGKGGDTVELTCTASQKSIQFHWNKSNQIK 59
QY 61 ILNGSGFLLTKGPSKLNDRADSRSLMDQGNFPLIINLKIEDSDTYICEVEDQKEEVQL 120
DB 60 ILNGSGFLLTKGPSKLNDRADSRSLMDQGNFPLIINLKIEDSDTYICEVEDQKEEVQL 117

```

```

Qy 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDG 180
Db 118 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPVQCRSPR-KNI-QGKTLVS-LELDG 174
Qy 181 TWCTCTVLQNKQKVEFKIDIVLAFOKASSIYKKEGEVSEFPLAFTVEKLTSGGLMW 240
Db 175 TWCTCTVLQNKQKVEFKIDIVLAFA-KASSIYKKEGEVSEFPLAFTVEKLTGS-ELMW 231
Qy 241 QAERASSSKSWITFDLKNKEVSVKRVYODPKLQNGKKLPLHLTLPOALPOYAGSGNLTIA 300
Db 232 QAERASSSKSWITFDLKNKEVSVKRVYODPKLQNGKKLPLHLTLPOALPOYAGSGNLTIA 291
Qy 301 LEAKTGKLGHEVNLVVMKATQLOKNTLCEVWGPTS PKLMLSLKLENKAKVSKKEKPYWV 360
Db 292 LEAKTGKLGHEVNLVVMKATQLOKNTLCEVWGPTS PKLMLSLKLENKAKVSKKEKPYWV 351
Qy 361 LNPEAGMWQCLISDSGVLLSNTKVLFTWSTPVHPRASALPAPPTGSALPDPTASALP 420
Db 352 LNPEAGMWQCLISDSGVLLSNTKVLFTWSTPVQP----- 387
Qy 421 DPAASALPALAVISFLGLGV-ACVLARTR 453
Db 388 -----MALIVGVAAGLLPLTGLGFFCVRGRHR 416

```

## RESULT 21

```

US-08-485-163-7
; Sequence 7, Application US/08485163
; Publication No. US20020098191A1
; GENERAL INFORMATION:
; APPLICANT: Beaudry, Gary A.
; TITLE OF INVENTION: CD4-GAMMA2 CD4-IgG2 CHIMERAS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,163
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 37690-II-1-ECT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; TELEX:
; INFORMATION FOR SEQ ID NO. 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 310 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: homo sapien
; CELL TYPE: lymphocyte
; US-08-485-163-7

```

```

Query Match 45.2%; Score 1050.5; DB 8; Length 310;
Best Local Similarity 75.4%; Pred. No. 4.4e-73;
Matches 221; Conservative 12; Mismatches 41; Indels 19; Gaps 4;

```

```

Qy 1 MNRGVPFRHLLVLOALALPAATQGNKVYLGKKGDTVELCTASQKKSIOFHMKNNOIK 60
Db 1 MNRGVPFRHLLVLOALALPAATQGNKVYLGKKGDTVELCTASQKKSIOFHMKNNOIK 60
Qy 61 ILNGQSFLLTKGSKLNDRASSRSIMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVOL 120
Db 61 ILNGQSFLLTKGSKLNDRASSRSIMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVOL 120
Qy 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDG 180
Db 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDG 180
Qy 181 TWCTCTVLQNKQKVEFKIDIVLAFOKA--SSIYKKEGEVSEFPLAFTVEKLTG---- 234
Db 181 TWCTCTVLQNKQKVEFKIDIVLAFTVAAPSVFFPPSEDEQLKSG--TASVYCLLNPFPR 238
Qy 235 SGEILMQAERASSSKSWITFDLKNKEVSVKRVYODPKLQNGKKLPLHLTLPOA 287
Db 239 EAKVQMKVDNALQSGN-----SQESVTEQDSKDSSTYLSSTLTLSKA 280

```

## RESULT 22

```

US-09-766-995-6
; Sequence 6, Application US/09766995
; Patent No. US20020052481A1
; GENERAL INFORMATION:
; APPLICANT: Graham P. Allaway et al.
; TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED CD4-GAMMA2 AND CD4-IgG2 IMMUNOCONJ
; FILE REFERENCE: 2048/41215-CB/JPW/SHS
; CURRENT APPLICATION NUMBER: US/09/766,995
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 310
; TYPE: PRT
; ORGANISM: homo sapiens
; US-09-766-995-6

```

```

Query Match 45.2%; Score 1050.5; DB 9; Length 310;
Best Local Similarity 75.4%; Pred. No. 4.4e-73;
Matches 221; Conservative 12; Mismatches 41; Indels 19; Gaps 4;

```

```

Qy 1 MNRGVPFRHLLVLOALALPAATQGNKVYLGKKGDTVELCTASQKKSIOFHMKNNOIK 60
Db 1 MNRGVPFRHLLVLOALALPAATQGNKVYLGKKGDTVELCTASQKKSIOFHMKNNOIK 60
Qy 61 ILNGQSFLLTKGSKLNDRASSRSIMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVOL 120
Db 61 ILNGQSFLLTKGSKLNDRASSRSIMDQGNFPLIINKLTIEDSDTYICEVEDQKEEVOL 120
Qy 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDG 180
Db 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELDG 180
Qy 181 TWCTCTVLQNKQKVEFKIDIVLAFOKA--SSIYKKEGEVSEFPLAFTVEKLTG---- 234
Db 181 TWCTCTVLQNKQKVEFKIDIVLAFTVAAPSVFFPPSEDEQLKSG--TASVYCLLNPFPR 238
Qy 235 SGEILMQAERASSSKSWITFDLKNKEVSVKRVYODPKLQNGKKLPLHLTLPOA 287
Db 239 EAKVQMKVDNALQSGN-----SQESVTEQDSKDSSTYLSSTLTLSKA 280

```

## RESULT 23

```

US-08-485-163-5
; Sequence 5, Application US/08485163
; Publication No. US20020098191A1
; GENERAL INFORMATION:
; APPLICANT: Beaudry, Gary A.
; US-08-485-163-5

```

```

; TITLE OF INVENTION: CD4-GAMMA2 CD4-IgG2 CHIMERAS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,163
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 37690-II-1-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: homo sapien
; CELL TYPE: lymphocyte
; US-08-485-163-5

Query Match      45.0%; Score 1046; DB 8; Length 530;
Best Local Similarity 93.4%; Pred. No. 2e-72;
Matches 211; Conservative 2; Mismatches 3; Indels 10; Gaps 3;

QY      1  MNRGVPFRHLILVLTQALLPATQGNKVLGKGGDTVELTCTASQKSIQPHMKNNOIK 60
DB      1  MNRGVPFRHLILVLTQALLPATQGNKVLGKGGDTVELTCTASQKSIQPHMKNNOIK 60
QY      61  ILNGGSFLTGGPSKLNDRADSRSLMDQGNFPLIIKNLKIETSDTYICEVEDQKEEYQL 120
DB      61  ILNGGSFLTGGPSKLNDRADSRSLMDQGNFPLIIKNLKIETSDTYICEVEDQKEEYQL 120
QY      121 LVFGLTANSDFHLLOGOSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB      121 LVFGLTANSDFHLLOGOSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY      121 LVFGLTANSDFHLLOGOSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB      121 LVFGLTANSDFHLLOGOSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY      181 TWCTVLONQKKEFKIDIVLAFQKASSIVYKKEGEQVEFSPFLA 226
DB      181 TWCTVLONQKKEFKIDIVLAFQKASSIVYKKEGEQVEFSPFLA 226
QY      181 TWCTVLONQKKEFKIDIVLAFQKASSIVYKKEGEQVEFSPFLA 216
DB      181 TWCTVLONQKKEFKIDIVLAFQKASSIVYKKEGEQVEFSPFLA 216

RESULT 24
; US-09-766-995-4
; Sequence 4, Application US/09766995
; Patent No. US2002052481A1
; GENERAL INFORMATION:
; APPLICANT: Graham P. Allaway et al.
; TITLE OF INVENTION: NON-PEPTIDYL MOLETY-CONJUGATED CD4-GAMMA2 AND CD4-IgG2 IMMUNOCONJ
; FILE REFERENCE: 2048/41215-CB/JPM/SHS
; CURRENT APPLICATION NUMBER: US/09/766,995
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
```

```

; LENGTH: 530
; TYPE: PRT
; ORGANISM: homo sapiens
; US-09-766-995-4

Query Match      45.0%; Score 1046; DB 9; Length 530;
Best Local Similarity 93.4%; Pred. No. 2e-72;
Matches 211; Conservative 2; Mismatches 3; Indels 10; Gaps 3;

QY      1  MNRGVPFRHLILVLTQALLPATQGNKVLGKGGDTVELTCTASQKSIQPHMKNNOIK 60
DB      1  MNRGVPFRHLILVLTQALLPATQGNKVLGKGGDTVELTCTASQKSIQPHMKNNOIK 60
QY      61  ILNGGSFLTGGPSKLNDRADSRSLMDQGNFPLIIKNLKIETSDTYICEVEDQKEEYQL 120
DB      61  ILNGGSFLTGGPSKLNDRADSRSLMDQGNFPLIIKNLKIETSDTYICEVEDQKEEYQL 120
QY      121 LVFGLTANSDFHLLOGOSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB      121 LVFGLTANSDFHLLOGOSITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
QY      181 TWCTVLONQKKEFKIDIVLAFQKASSIVYKKEGEQVEFSPFLA 226
DB      181 TWCTVLONQKKEFKIDIVLAFQKASSIVYKKEGEQVEFSPFLA 216

RESULT 25
; US-08-485-163-3
; Sequence 3, Application US/08485163
; Publication No. US20020098191A1
; GENERAL INFORMATION:
; APPLICANT: Beaudry, Gary A.
; APPLICANT: Maddon, Paul J.
; TITLE OF INVENTION: CD4-GAMMA2 CD4-IgG2 CHIMERAS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,163
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 37690-II-1-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 432 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: homo sapien
; CELL TYPE: lymphocyte
; US-08-485-163-3

Query Match      44.9%; Score 1045; DB 8; Length 432;
Best Local Similarity 98.5%; Pred. No. 1.8e-72;
```



```

; TITLE OF INVENTION: of Use
; FILE REFERENCE: P-IS 5155
; CURRENT APPLICATION NUMBER: US/10/125,692
; CURRENT FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: US 60/285,477
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 612
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-692-10

Query Match      42.7%; Score 993; DB 14; Length 612;
Best Local Similarity 53.8%; Pred. No. 3.2e-68;
Matches 211; Conservative 62; Mismatches 109; Indels 10; Gaps 7;

QY 1 MNRGVPRH-LTLVQLALPPATQGNKVLGKGGDTVELCTASQKKSIOFHWKNSNOI 59
DB 1 MCRALSRLRLLLQLSGLAVTQKTLVLGKEGESALPCESSQKITYFTWKFSQOR 60
QY 60 KILNQG-SFLTKG--PSKLRADRSRLMDQGNFPLIIKNLKIEDSDTYICEVEDQKE 116
DB 61 KILGKHGKGVLRGSSPSQF-DRFDSKKGAMEKGFPIIKNLKNEDSQTYICELENKE 119
QY 117 EVQLLVFGLTANSDPHLLQGSLITLES-PPGSSPSVQCRSPRGKNIQGGKTLVSQLE 175
DB 120 EVELVVFVKTSPGTSLLQGOSLTITLDSNSKVSNPLETECKKKKVVSGSVLSMSNLR 179
QY 176 LODSGTWTCTVLOQNKVEFKIDIVLAFQKASIVYKKEGQVSEFPPLAFTVEKLGS 235
DB 180 VODSPFNCTVTLDDQKMFQMTLVLGFSQTAITAYISSEGSASFSPPLNFAER--NGW 237
QY 236 GELMWQAERASSSKSMITFDLKNKEVSKRYTQDPKLOMKKLPILHTLPQALPOYASG 295
DB 238 GELMWKAERKDSFQFQWISFISIKKEVSQKSTKDLQKERTPLTLKIPQVSLQFASG 297
QY 236 NLTLLAEKTKLHDEVNLVYMRATQLOKNTLCEVWGPSPSLMLSLLENKAVSKRE 355
DB 238 NLTLLD--KGTLHDEVNLVYMKVQALNNLTLCCEVWGPSPKMRILTLKQENQEARVSEQ 355
QY 356 KPVVTLNPEAGMMOCLLSDSGVLLSESNIKVL 387
DB 356 KVVQVAPETGLMOCLLSEGDKVKMDSRIQVL 387

RESULT 29
US-10-073-118-26
; Sequence 26, Application US/10073118
; Publication No. US2003005454A1
; GENERAL INFORMATION:
; APPLICANT: BECQUART, JEROME
; TITLE OF INVENTION: ALBUMIN DERIVATIVES WITH THERAPEUTIC FUNCTIONS
; FILE REFERENCE: 06832.1429-03
; CURRENT FILING DATE: 2002-02-12
; PRIOR APPLICATION NUMBER: US/10/073,118
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: 09/551,635
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: 09/004,319
; PRIOR FILING DATE: 1998-01-08
; PRIOR APPLICATION NUMBER: 08/479,146
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/295,078
; PRIOR FILING DATE: 1994-08-26
; PRIOR APPLICATION NUMBER: 08/121,236
; PRIOR FILING DATE: 1993-09-13
; PRIOR APPLICATION NUMBER: 07/955,243
; PRIOR FILING DATE: 1992-10-01
; PRIOR APPLICATION NUMBER: 07/561,879
; PRIOR FILING DATE: 1990-08-02
; PRIOR APPLICATION NUMBER: FR 89 10480
; PRIOR FILING DATE: 1989-08-03
```

```

; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 788
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Protein fusion
US-10-073-118-26

Query Match      39.6%; Score 921; DB 14; Length 788;
Best Local Similarity 95.8%; Pred. No. 1.7e-62;
Matches 182; Conservative 1; Mismatches 3; Indels 4; Gaps 1;

QY 19 LPATQG---NKVVLGKGDVLTCTASQKKSIOFHWKNSNOIKILNQGSFLTKGPS 74
DB 599 LVAAQALGLKTKVTLGKGDVLTCTASQKKSIOFHWKNSNOIKILNQGSFLTKGPS 658
QY 75 KLRNADRSRLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDPHLL 134
DB 659 KLRNADRSRLMDQGNFPLIIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDPHLL 718
QY 135 QGOSLTITLESPPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTWTCTVLOQNKVE 194
DB 719 QGOSLTITLESPPSSPSVQCRSPRGKNIQGGKTLVSQLELQDSGTWTCTVLOQNKVE 778
QY 195 FKIDIVLAF 204
DB 779 FKIDIVLAF 788

RESULT 30
US-09-934-060A-13
; Sequence 13, Application US/09934060A
; Patent No. US2002015121A1
; GENERAL INFORMATION:
; APPLICANT: Devico, Anthony L.
; APPLICANT: Fouts, Timothy R.
; TITLE OF INVENTION: VIRUS COAT PROTEIN/RECEPTOR CHIMERAS AND METHODS OF USE
; FILE REFERENCE: 4115-144 CIP
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US/09/934,060A
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 09/684,026
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/158,321
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 590
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthesized construct
; NAME/KEY: MISC FEATURE
; LOCATION: (586)..(586)
; OTHER INFORMATION: Xaa can be any amino acid
; NAME/KEY: MISC FEATURE
; LOCATION: (589)..(589)
; OTHER INFORMATION: Xaa can be any amino acid
US-09-934-060A-13

Query Match      39.5%; Score 919; DB 9; Length 590;
Best Local Similarity 98.4%; Pred. No. 1.7e-62;
Matches 180; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 27 KVLIGKGDVLTCTASQKKSIOFHWKNSNOIKILNQGSFLTKGPSKLRNADRSRL 86
DB 399 KVLIGKGDVLTCTASQKKSIOFHWKNSNOIKILNQGSFLTKGPSKLRNADRSRL 458
QY 87 MDQGNFPLIIKNLKIEDSDTYICEVEDQKEVQLVFGLTANSDPHLLQGOSLTITLES 146
```

Db 459 WDQGNFLLIKNLKIEISDPTICEVEEDQKEEVOLLVFGLTANSSTHLLQGOSLTLTLESP 518  
Qy 147 PGSSPSVQCSRPGRKNIQGKTLVSQLELDSDGTMCTVYLQNKVPEFKIDIVVLAFOK 206  
Db 519 PGSSPSVQCSRPGRKNIQGKTLVSQLELDSDGTMCTVYLQNKVPEFKIDIVVLAFOK 578  
Qy 207 ASS 209  
Db 579 LIS 581

RESULT 31  
US-09-934-060A-2  
; Sequence 2, Application US/09934060A  
; Patent No. US20020155121A1  
; GENERAL INFORMATION:  
; APPLICANT: Devico, Anthony L.  
; APPLICANT: Fouts, Timothy R.  
; APPLICANT: Tuskan, Robert G.  
; TITLE OF INVENTION: VIRUS COAT PROTEIN/RECEPTOR CHIMERAS AND METHODS OF USE  
; FILE REFERENCE: 4115-144 CIP  
; CURRENT APPLICATION NUMBER: US/09/934,060A  
; CURRENT FILING DATE: 2001-08-21  
; PRIOR APPLICATION NUMBER: US 09/684,026  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/158,321  
; PRIOR FILING DATE: 1999-10-08  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 720  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (716)..(716)  
; OTHER INFORMATION: Xaa can be any amino acid  
; NAME/KEY: MISC FEATURE  
; LOCATION: (719)..(719)  
; OTHER INFORMATION: Xaa can be any amino acid  
US-09-934-060A-2

Query Match 39.5%; Score 919; DB 9; Length 720;  
Best Local Similarity 98.4%; Pred. No. 2,2e-62;  
Matches 180; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 27 KVLGKGGDVELTCTASQKKSIOFHWKNSNQIKILGQGSFLTGPSTKINDRADSRSL 86  
Db 529 KVLGKGGDVELTCTASQKKSIOFHWKNSNQIKILGQGSFLTGPSTKINDRADSRSL 588  
Qy 87 WDQGNFLLIKNLKIEISDPTICEVEEDQKEEVOLLVFGLTANSSTHLLQGOSLTLTLESP 146  
Db 589 WDQGNFLLIKNLKIEISDPTICEVEEDQKEEVOLLVFGLTANSSTHLLQGOSLTLTLESP 648  
Qy 147 PGSSPSVQCSRPGRKNIQGKTLVSQLELDSDGTMCTVYLQNKVPEFKIDIVVLAFOK 206  
Db 649 PGSSPSVQCSRPGRKNIQGKTLVSQLELDSDGTMCTVYLQNKVPEFKIDIVVLAFOK 708  
Qy 207 ASS 209  
Db 709 LIS 711

RESULT 32  
US-09-934-060A-4  
; Sequence 4, Application US/09934060A  
; Patent No. US20020155121A1  
; GENERAL INFORMATION:  
; APPLICANT: Devico, Anthony L.  
; APPLICANT: Fouts, Timothy R.  
; APPLICANT: Tuskan, Robert G.

; TITLE OF INVENTION: VIRUS COAT PROTEIN/RECEPTOR CHIMERAS AND METHODS OF USE  
; FILE REFERENCE: 4115-144 CIP  
; CURRENT APPLICATION NUMBER: US/09/934,060A  
; CURRENT FILING DATE: 2001-08-21  
; PRIOR APPLICATION NUMBER: US 09/684,026  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/158,321  
; PRIOR FILING DATE: 1999-10-08  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 720  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (716)..(716)  
; OTHER INFORMATION: Xaa can be any amino acid  
; NAME/KEY: MISC FEATURE  
; LOCATION: (719)..(719)  
; OTHER INFORMATION: Xaa can be any amino acid  
US-09-934-060A-4

Query Match 39.5%; Score 919; DB 9; Length 720;  
Best Local Similarity 98.4%; Pred. No. 2,2e-62;  
Matches 180; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 27 KVLGKGGDVELTCTASQKKSIOFHWKNSNQIKILGQGSFLTGPSTKINDRADSRSL 86  
Db 529 KVLGKGGDVELTCTASQKKSIOFHWKNSNQIKILGQGSFLTGPSTKINDRADSRSL 588  
Qy 87 WDQGNFLLIKNLKIEISDPTICEVEEDQKEEVOLLVFGLTANSSTHLLQGOSLTLTLESP 146  
Db 589 WDQGNFLLIKNLKIEISDPTICEVEEDQKEEVOLLVFGLTANSSTHLLQGOSLTLTLESP 648  
Qy 147 PGSSPSVQCSRPGRKNIQGKTLVSQLELDSDGTMCTVYLQNKVPEFKIDIVVLAFOK 206  
Db 649 PGSSPSVQCSRPGRKNIQGKTLVSQLELDSDGTMCTVYLQNKVPEFKIDIVVLAFOK 708  
Qy 207 ASS 209  
Db 709 LIS 711

RESULT 33  
US-09-934-060A-26  
; Sequence 26, Application US/09934060A  
; Patent No. US20020155121A1  
; GENERAL INFORMATION:  
; APPLICANT: Devico, Anthony L.  
; APPLICANT: Fouts, Timothy R.  
; APPLICANT: Tuskan, Robert G.  
; TITLE OF INVENTION: VIRUS COAT PROTEIN/RECEPTOR CHIMERAS AND METHODS OF USE  
; FILE REFERENCE: 4115-144 CIP  
; CURRENT APPLICATION NUMBER: US/09/934,060A  
; CURRENT FILING DATE: 2001-08-21  
; PRIOR APPLICATION NUMBER: US 09/684,026  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/158,321  
; PRIOR FILING DATE: 1999-10-08  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 26  
; LENGTH: 178  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthesized construct  
US-09-934-060A-26

Query Match 39.2%; Score 911; DB 9; Length 178;  
Best Local Similarity 100.0%; Pred. No. 1,3e-62;



Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 KVVLEKGGDTVELTCTASQKSIQPHMKNNOIKILGNQSFLLTGKPSKLNDRADSRSL 86  
DB 2 KVVLEKGGDTVELTCTASQKSIQPHMKNNOIKILGNQSFLLTGKPSKLNDRADSRSL 61

QY 87 MDQGNFLLIKNLKIEDSDTYICEVEDQKEEVQLVFGILTANSDPHLLQGGSLTLLTSSP 146  
DB 62 MDQGNFLLIKNLKIEDSDTYICEVEDQKEEVQLVFGILTANSDPHLLQGGSLTLLTSSP 121

QY 147 PGSSPSVOCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLOQKKEFKIDIVVLA 203  
DB 122 PGSSPSVOCRSPRGKNIQGGKTLVSQLELQDSGTWCTVLOQKKEFKIDIVVLA 178

RESULT 34  
US-10-024-329-33  
; Sequence 33, Application US/10024329  
; Publication No. US20030157063A1  
; GENERAL INFORMATION:  
; APPLICANT: SANHADJI, Kamel  
; APPLICANT: TOURAINE, Jean-Louis  
; APPLICANT: LEROY, Pierre  
; APPLICANT: MEHTALI, Majid  
; TITLE OF INVENTION: Gene therapy using anti-gp41 antibody and cd4 immunoadhesin  
; FILE REFERENCE: 109993  
; CURRENT APPLICATION NUMBER: US/10/024,329  
; CURRENT FILING DATE: 2001-12-21  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 33  
; LENGTH: 184  
; TYPE: PRT  
; ORGANISM: L, D1, J1, D2 domains of human SCDA  
US-10-024-329-33

Query Match 37.9%; Score 881; DB 14; Length 184;  
Best Local Similarity 95.3%; Pred. No. 3e-60; 2; Indels 6; Gaps 5;  
Matches 181; Conservative 1; Mismatches 2;

QY 1 MNRGVPRHLTLVQLALPAATQGNKVVVGKGGDTVELTCTASQKSIQPHMKNNOIK 60  
DB 1 MNRGVPR-HLLVLTQLALPAATQGNKVVVGKGGDTVELTCTASQKSIQPHMKNNOIK 59

QY 61 ILGNQSFLLTGKPSKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEEVQL 120  
DB 60 ILGNQSFLLTGKPSKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEEVQL 117

QY 121 LVFGILTANSDPHLLQGGSLTLLTSSPSSPSVOCRSPRGKNIQGGKTLVSQLELQDSG 180  
DB 118 LVFGILTANSDPHLLQGGSLTLLTSSPSSPSVOCRSPRGKNIQGGKTLVSQLELQDSG 174

QY 181 TWTCTVLOQK 190  
DB 175 TWTCTVLOQK 184

RESULT 35  
US-09-891-119A-10  
; Sequence 10, Application US/09891119A  
; Publication No. US20040013663A1  
; GENERAL INFORMATION:  
; APPLICANT: Maddon, Paul J.  
; TITLE OF INVENTION: DERIVATIVES OF SOLUBLE T-4  
; FILE REFERENCE: 24577-CY-B  
; CURRENT APPLICATION NUMBER: US/09/891,119A  
; CURRENT FILING DATE: 2001-06-25  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 94  
; TYPE: PRT  
; ORGANISM: human

US-09-891-119A-10

Query Match 21.2%; Score 494; DB 11; Length 94;  
Best Local Similarity 100.0%; Pred. No. 1.2e-30;  
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 QGNKVVLEKGGDTVELTCTASQKSIQPHMKNNOIKILGNQSFLLTGKPSKLNDRADSR 83  
DB 1 QGNKVVLEKGGDTVELTCTASQKSIQPHMKNNOIKILGNQSFLLTGKPSKLNDRADSR 60

QY 84 RSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEE 117  
DB 61 RSLMDQGNFLLIKNLKIEDSDTYICEVEDQKEE 94

RESULT 36  
US-10-105-545-26  
; Sequence 26, Application US/10105545  
; Publication No. US20030144479A1  
; GENERAL INFORMATION:  
; APPLICANT: Mark, Greene I.  
; APPLICANT: Williams, William V.  
; APPLICANT: Weiner, David B.  
; APPLICANT: Cohen, Jeffrey A.  
; APPLICANT: Kieber-Emmons, Thomas  
; APPLICANT: Williams, Robert M.  
; TITLE OF INVENTION: BIOLOGICALLY ACTIVE COMPOUNDS AND METHODS OF CONSTRUCTING AND USE  
; FILE REFERENCE: 4040/11492US2  
; CURRENT APPLICATION NUMBER: US/10/105,545  
; CURRENT FILING DATE: 2002-09-06  
; PRIOR APPLICATION NUMBER: US 08/752,816  
; PRIOR FILING DATE: 1996-11-21  
; PRIOR APPLICATION NUMBER: US 07/940,654  
; PRIOR FILING DATE: 1992-09-03  
; PRIOR APPLICATION NUMBER: US 07/702,833  
; PRIOR FILING DATE: 1991-05-20  
; PRIOR APPLICATION NUMBER: US 07/326,328  
; PRIOR FILING DATE: 1989-03-21  
; PRIOR APPLICATION NUMBER: US 07/074,264  
; PRIOR FILING DATE: 1987-07-16  
; PRIOR APPLICATION NUMBER: US 07/462,542  
; PRIOR FILING DATE: 1990-01-09  
; PRIOR APPLICATION NUMBER: US 07/648,303  
; PRIOR FILING DATE: 1991-01-25  
; PRIOR APPLICATION NUMBER: US 07/685,881  
; PRIOR FILING DATE: 1991-04-15  
; PRIOR APPLICATION NUMBER: US 07/574,391  
; PRIOR FILING DATE: 1990-08-27  
; PRIOR APPLICATION NUMBER: US 07/194,026  
; PRIOR FILING DATE: 1988-05-13  
; Remaining Prior Application data removed - See file wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 26  
; LENGTH: 93  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetic peptide  
US-10-105-545-26

Query Match 20.9%; Score 486; DB 14; Length 93;  
Best Local Similarity 100.0%; Pred. No. 4.9e-30;  
Matches 93; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALPPATQGNKVVVGKGGDTVELTCTASQKSIQPHMKNNOIKILGNQSFLLTGKPSK 76  
DB 1 ALPPATQGNKVVVGKGGDTVELTCTASQKSIQPHMKNNOIKILGNQSFLLTGKPSK 60

QY 77 NDRADSRSLMDQGNFLLIKNLKIEDSDTYIC 109  
DB 61 NDRADSRSLMDQGNFLLIKNLKIEDSDTYIC 93

RESULT 37  
US-09-929-924-16  
; Sequence 16, Application US/09929924  
; Patent No. US20020147151A1  
; GENERAL INFORMATION:  
; APPLICANT: Center, David M.  
; APPLICANT: Cruickshank, William W.  
; APPLICANT: Kornfeld, Hardy  
; TITLE OF INVENTION: IL-16 ANTAGONISTS  
; FILE REFERENCE: Research Corporation Tech., Inc.  
; CURRENT APPLICATION NUMBER: US/09/929,924  
; CURRENT FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: 09/368,632  
; PRIOR FILING DATE: 1999-08-05  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 73  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: IL-16 antagonist peptide  
US-09-929-924-16

Query Match 16.0%; Score 373; DB 9; Length 73;  
Best Local Similarity 97.2%; Pred. No. 2e-21;  
Matches 70; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 325 NLTCVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPAAGMOCILSDSGQVLLSNI 384  
DB 1 NLTCVWGPTSPKMLSLKLENKEAKVSKREKAVWVLPNPAAGMOCILSDSGQVLLSNI 60

QY 385 KVLPTWSTPVHP 396  
DB 61 KVLPTWSTPVOP 72

RESULT 38  
US-09-929-924-36  
; Sequence 36, Application US/09929924  
; Patent No. US20020147151A1  
; GENERAL INFORMATION:  
; APPLICANT: Center, David M.  
; APPLICANT: Cruickshank, William W.  
; APPLICANT: Kornfeld, Hardy  
; TITLE OF INVENTION: IL-16 ANTAGONISTS  
; FILE REFERENCE: Research Corporation Tech., Inc.  
; CURRENT APPLICATION NUMBER: US/09/929,924  
; CURRENT FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: 09/368,632  
; PRIOR FILING DATE: 1999-08-05  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 36  
; LENGTH: 73  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: IL-16 antagonist peptide  
US-09-929-924-36

Query Match 16.0%; Score 371; DB 9; Length 73;  
Best Local Similarity 95.8%; Pred. No. 2.9e-21;  
Matches 69; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 325 NLTCVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPAAGMOCILSDSGQVLLSNI 384  
DB 1 NLTCVWGPTSPKMLSLKLENKEAKVSKREKAVWVLPNPAAGMOCILSDSGQVLLSNI 60

QY 385 KVLPTWSTPVHP 396  
DB 61 KVLPTWSTPVOP 72

DB 61 KVLPTWSTPVOP 72

RESULT 39  
US-09-929-924-37  
; Sequence 37, Application US/09929924  
; Patent No. US20020147151A1  
; GENERAL INFORMATION:  
; APPLICANT: Center, David M.  
; APPLICANT: Cruickshank, William W.  
; APPLICANT: Kornfeld, Hardy  
; TITLE OF INVENTION: IL-16 ANTAGONISTS  
; FILE REFERENCE: Research Corporation Tech., Inc.  
; CURRENT APPLICATION NUMBER: US/09/929,924  
; CURRENT FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: 09/368,632  
; PRIOR FILING DATE: 1999-08-05  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 37  
; LENGTH: 73  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: IL-16 antagonist peptide  
US-09-929-924-37

Query Match 16.0%; Score 371; DB 9; Length 73;  
Best Local Similarity 95.8%; Pred. No. 2.9e-21;  
Matches 69; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 325 NLTCVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPAAGMOCILSDSGQVLLSNI 384  
DB 1 NLTCVWGPTSPKMLSLKLENKEAKVSKREKAVWVLPNPAAGMOCILSDSGQVLLSNI 60

QY 385 KVLPTWSTPVHP 396  
DB 61 KVLPTWSTPVOP 72

RESULT 40  
US-09-929-924-33  
; Sequence 33, Application US/09929924  
; Patent No. US20020147151A1  
; GENERAL INFORMATION:  
; APPLICANT: Center, David M.  
; APPLICANT: Cruickshank, William W.  
; APPLICANT: Kornfeld, Hardy  
; TITLE OF INVENTION: IL-16 ANTAGONISTS  
; FILE REFERENCE: Research Corporation Tech., Inc.  
; CURRENT APPLICATION NUMBER: US/09/929,924  
; CURRENT FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: 09/368,632  
; PRIOR FILING DATE: 1999-08-05  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 33  
; LENGTH: 73  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: IL-16 antagonist peptide  
US-09-929-924-33

Query Match 15.8%; Score 368; DB 9; Length 73;  
Best Local Similarity 95.8%; Pred. No. 4.9e-21;  
Matches 69; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 325 NLTCVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPAAGMOCILSDSGQVLLSNI 384  
DB 1 NLTCVWGPTSPKMLSLKLENKEAKVSKREKAVWVLPNPAAGMOCILSDSGQVLLSNI 60

QY 385 KVLPTWSTPVHP 396  
DB 61 KVLPTWSTPVOP 72

Db 61 KVLPTMSTPVP 72

## RESULT 41

US-09-929-924-34  
; Sequence 34, Application US/09929924  
; Patent No. US20020147151A1  
; GENERAL INFORMATION:  
; APPLICANT: Center, David M.  
; APPLICANT: Cruickshank, William W.  
; APPLICANT: Kornfeld, Hardy  
; TITLE OF INVENTION: IL-16 ANTAGONISTS  
; FILE REFERENCE: Research Corporation Tech., Inc.  
; CURRENT APPLICATION NUMBER: US/09/929,924  
; CURRENT FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: 09/368,632  
; PRIOR FILING DATE: 1999-08-05  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 34  
; LENGTH: 73  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: IL-16 antagonist peptide  
US-09-929-924-34

Query Match 15.8%; Score 367; DB 9; Length 73;  
Best Local Similarity 95.8%; Pred. No. 5.9e-21;  
Matches 69; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 335 NITCEVWGPTSPKMLSLKENKAVSKREKPVAVLNPEAGMOCCLSDSGVLLSNI 384  
Db 1 NITCEVWGPTSPKMLSLKENKAVSKREKPVAVLNPEAGMOCCLSDSGVLLSNI 60  
Qy 385 KVLPTMSTPVP 396  
Db 61 KVLPTMSTPVP 72

## RESULT 42

US-09-929-924-35  
; Sequence 35, Application US/09929924  
; Patent No. US20020147151A1  
; GENERAL INFORMATION:  
; APPLICANT: Center, David M.  
; APPLICANT: Cruickshank, William W.  
; APPLICANT: Kornfeld, Hardy  
; TITLE OF INVENTION: IL-16 ANTAGONISTS  
; FILE REFERENCE: Research Corporation Tech., Inc.  
; CURRENT APPLICATION NUMBER: US/09/929,924  
; CURRENT FILING DATE: 2001-08-15  
; PRIOR APPLICATION NUMBER: 09/368,632  
; PRIOR FILING DATE: 1999-08-05  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 35  
; LENGTH: 73  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: IL-16 antagonist peptide  
US-09-929-924-35

Query Match 15.8%; Score 367; DB 9; Length 73;  
Best Local Similarity 95.8%; Pred. No. 5.9e-21;  
Matches 69; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 335 NITCEVWGPTSPKMLSLKENKAVSKREKPVAVLNPEAGMOCCLSDSGVLLSNI 384  
Db 1 NITCEVWGPTSPKMLSLKENKAVSKREKPVAVLNPEAGMOCCLSDSGVLLSNI 60

Qy 385 KVLPTMSTPVP 396  
Db 61 KVLPTMSTPVP 72

## RESULT 43

US-09-135-238B-8  
; Sequence 8, Application US/09135238B  
; Patent No. US20020177565A1  
; GENERAL INFORMATION:  
; APPLICANT: No. US20020177565A1an, Garry P.  
; APPLICANT: Hitooshi, Yasumichi  
; TITLE OF INVENTION: TOSO  
; FILE REFERENCE: A65635-1/DJB/RMS  
; CURRENT APPLICATION NUMBER: US/09/135,238B  
; CURRENT FILING DATE: 1998-08-17  
; PRIOR APPLICATION NUMBER: 60/066,063  
; PRIOR FILING DATE: 1997-11-17  
; NUMBER OF SEQ ID NOS: 31  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 84  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (11)..(14)  
; OTHER INFORMATION: The xaa at positions 11 through 14 represents an  
; OTHER INFORMATION: unknown amino acid.  
; NAME/KEY: UNSURE  
; LOCATION: (28)  
; OTHER INFORMATION: The xaa at position 28 represents an unknown amino  
; OTHER INFORMATION: acid.  
; NAME/KEY: UNSURE  
; LOCATION: (41)  
; OTHER INFORMATION: The xaa at position 41 represents an unknown amino  
; OTHER INFORMATION: acid.  
; NAME/KEY: UNSURE  
; LOCATION: (46)..(48)  
; OTHER INFORMATION: The xaa at positions 46 through 48 represents an  
; OTHER INFORMATION: unknown amino acid.  
; NAME/KEY: UNSURE  
; LOCATION: (62)  
; OTHER INFORMATION: The xaa at position 62 represents an unknown amino  
; OTHER INFORMATION: acid.  
US-09-135-238B-8

Query Match 14.5%; Score 337; DB 9; Length 84;  
Best Local Similarity 88.1%; Pred. No. 1.5e-18;  
Matches 74; Conservative 0; Mismatches 0; Indels 10; Gaps 5;

Qy 37 VELTCTASQK---KSIQFHKNSNQI-KIIGNGSFLTK-GPSK--LNDRADSRSLW 87  
Db 1 VELTCTASQKXXXXKSIQFHKNSNQIKIIGNGSFLTKGPSKXXLLNDRADSRSLW 60  
Qy 88 D-QGNFPLIKNLKIEDSDTYICE 110  
Db 61 DQGNFPLIKNLKIEDSDTYICE 84

## RESULT 44

US-09-997-165-2  
; Sequence 2, Application US/09997165  
; Patent No. US20020141999A1  
; GENERAL INFORMATION:  
; APPLICANT: Lyman, Stewart D.  
; APPLICANT: Fanslow, William C.  
; TITLE OF INVENTION: LIGAND FOR CD7 AND METHODS OF USE THEREOF  
; FILE REFERENCE: 2913-US  
; CURRENT APPLICATION NUMBER: US/09/997,165  
; CURRENT FILING DATE: 2001-11-27  
; PRIOR APPLICATION NUMBER: PCT/US00/14612  
; PRIOR FILING DATE: 2000-05-26





```

; CURRENT FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: PCT/US01/03651
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/632,085
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 586
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-311-823-7

Query Match      7.0%; Score 162; DB 16; Length 586;
Best Local Similarity 22.1%; Pred. No. 0.0008;
Matches 108; Conservative 77; Mismatches 185; Indels 118; Gaps 25;

Qy      4 GVPRHLLVQLLPLPATQGNVYLGKGD-----TVELTCTASQKSIQFHWKNSQ 58
Db      115 GVPIKIRVDVQYDEPMLTV-HQTVSDVRGNFYQEKTVFLRCTVNSNPAPRFTMKRGS 173
Qy      59 IKILNGQSFITKGPSTKLNDRADSRSLMDQGNFPLI-IKULKIEDSDTYICEVEDQKE- 116
Db      174 -----TSHSQDNG-VDIYEPLTYQGETKVLKLNLRQDYASTYCVSVANVC 221
Qy      117 --EVQLLVFGLTANSDFHLLQGQSLTTLTSSPGSSPSVOC-----R 156
Db      222 GIPDKATIFRLTNTTAPPAK-LSVNETLVNPGENVTQCLLTGSGDPLPQLQWSHGPGF 280
Qy      157 SPREKNIQGGKTLVSQLELDSDGTWCTVLYQN-----QKVEFKIDIVLAFQKASSIVY 212
Db      281 LPLDALAQGG-TLSIPVQARDSGYVNCATANNGNPAKKT---VNLVSRMKNATFQIT 336
Qy      213 K---KEGQVEFSFPL-----AFVTEKLTGSGELWMQ-AERASSKSWITFDLKKKEV 261
Db      337 PDVKESENIOLOQDGLKSCHVDAVPQEKVYQ---WFKNGKPARMSKRL--VTNDP 390
Qy      262 SVKRVTDPKLQMKGKLPYLH-----TLPOA-LPOYAGSGNLTALAEATGKLHQ 310
Db      391 ELPAVTSLSLEL-----IDLHFSYDGYTLCMASFGAPRDPDLSVEVNISSETVPTISVPK 445
Qy      311 EVNLVWNRATOLQNLTCGVWGPSTPKMLSLKLENKAK-----VSKREKPVVLANPE- 364
Db      446 GRAVVTVREGS-PAELOCEVARGKRPVPLWS--RVDKAALLPSGLPLEETPDGKRLER 502
Qy      365 -----AGMOCILSDSG-----QVLLSNIKV-----LPTWSTPVHPR 397
Db      503 VSRMSGTYRCQTRARYNGFNVRPREAOVQNLVQCESTTPGLPSQLSFWLPAIGSLPHPL 562
Qy      398 ASALPAPP 405
Db      563 THPSPTKP 570

RESULT 50
US-10-094-886-52
; Sequence 52, Application US/10094886
; Publication No. US20040002120A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Paturajan, Meera
; APPLICANT: Burgess, Catherine
; APPLICANT: Vernet, Corine A.
; APPLICANT: Li, Li
; APPLICANT: Gorman, Linda
; APPLICANT: Malyankar, Uriel M.
```

```

; APPLICANT: Boldog, Ferenc
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shenoy, Suresh
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Taupier, Raymond J., Jr.
; APPLICANT: Miller, Charles
; APPLICANT: Casman, Stacie
; APPLICANT: Pena, Carol
; APPLICANT: Gangolli, Reha
; APPLICANT: Gusev, Vladimir
; APPLICANT: Smltson, Glenda
; APPLICANT: Zehusen, Bryan
; APPLICANT: Gerlach, Valerie
; APPLICANT: Pochart, Pascal
; APPLICANT: Fernandes, Elma
; APPLICANT: Shinkets, Richard
; APPLICANT: Rastelli, Luca
; APPLICANT: Spaderna, Steven
; APPLICANT: Larochele, William
; APPLICANT: Zhong, Wei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS
; FILE REFERENCE: 21402-290 B
; CURRENT APPLICATION NUMBER: US/10/094,886
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/274,322
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/313,182
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/288,052
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/318,510
; PRIOR FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/314,018
; PRIOR FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/296,693
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: 60/313,626
; PRIOR FILING DATE: 2001-08-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 298
; SOFTWARE: PatentIn 2.1
; SEQ ID NO 52
; LENGTH: 1315
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-886-52

Query Match      6.8%; Score 157; DB 15; Length 1315;
Best Local Similarity 23.3%; Pred. No. 0.0059;
Matches 112; Conservative 64; Mismatches 181; Indels 124; Gaps 25;

Qy      36 TVELTCTASQKSIQFHWKNSQIKILNGQSGFLTGPSKLNDRADSRSLMDQGNFPLI 95
Db      152 TVFLRCTVNSNPAPRFTMKRGS-----TSHSQDNG-VDIYEPLTYQGETKVL 199
Qy      96 -IKULKIEDSDTYICEVEDQKE---EVQLLVFGLTANSDFHLLQGQSLTTLTSSPGSSP 151
Db      200 KLNLRPQDYASTYCVSVANVCIGIDPKATIFRLTNTTAPPAK-LSVNETLVNPGENV 258
Qy      152 SVQC-----RSPRKNIQGGKTLVSQLELDSDGTWCTVLYQN----- 189
Db      259 TVQCLLTGSDPLPQLQWSHGPGPLPGALAQGG-TLSIPVQARDSGYVNCATANNVGNP 317
Qy      190 QKVEFKIDIVLAFQKASSIVK---KEGQVEFSFPL-----AFVTEKLTGSGELW 239
Db      318 AKKT---VNLVSRMKNATFQITPDVTKESENIOLOQDGLKSCHVDAVPQEKVYQ---W 371
```



NUMBER OF SEQ ID NOS: 101  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 69  
LENGTH: 739  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-047-542-69

Query Match 6.7%; Score 155; DB 13; Length 739;  
Best Local Similarity 21.9%; Pred. No. 0.0038;  
Matches 91; Conservative 71; Mismatches 166; Indels 88; Gaps 18;

QY RHLVLVQLALPPAT--QGNKVVLGKKGVLTCTASQKKSIOFHWKNSNOIKILGNQ 65  
DB RQAVKEIQVIVSPKNTVISVNPSTKLOEGSVTWTCSSEGLPAPEIFW----- 258  
QY 66 GSFLTKGPSKLNDRASRLMDQGNFPLIKNLKIEDSDTYICE---VEDQKEEYQL 121  
DB 259 -----SKLIDNGNLQHL---SGNATLTLIAMRMEDSGIYVCEGVNLIGKRNKEVELI 307  
QY 259 -----SKLIDNGNLQHL---SGNATLTLIAMRMEDSGIYVCEGVNLIGKRNKEVELI 307  
DB 122 V----FGLTANSDTHLLQ--GQSLTLTLSPSSPSVQCRSPRGKNIQG-----GKT 168  
DB 308 VOEKPFTVEISPGPRIAQAIGDSVMLTCSVMGCESPFSFWRTQIDPSLGSKVSEGTNST 367  
QY 169 LSVSQAELQDSGWTCTVLAQNKVFEKIDIVLAFOKASSIYK---KEGEQVEFS--- 222  
DB 368 LTLSPVSEFENHSHYLCVTGCKHKKELQVELYSPRPDEIEMSGLVNGSSVTYVSCV 427  
QY 223 ---PFL-AFTVEKLTSGSELMWQERASSKSWITF---DLKKEVSVKRVTDPKLQ- 273  
DB 428 PSYVPLDRLEIELKGETIL-----ENIEFLIEDTDMKSLKMSLEMTPIPTIED 476  
QY 274 MGKLL-----PLHLTLPOALPOYAGSGNLTALBAKTKLHOEVNLVYMRATOLQK---N 325  
DB 477 TKGALVQOAKLHIDMEFEKQKQOS---TQTLVYVNAV--RDTVLVSPSSILEGSSVN 531  
QY 326 LTCSEWGPSTPKMLSLKENKAKVSKREKPVWVLPNPEAGMOCCLSDSGVLL 381  
DB 532 MTLCSQGFAPAKILMSRQLPNEGELQISENATLTLISTK-----MEDSGVLYCE 580

RESULT 53  
US-10-207-655-139  
Sequence 139, Application US/10207655  
Publication No. US20030118592A1  
GENERAL INFORMATION:  
APPLICANT: Ledbetter, Jeffrey A.  
TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS  
FILE REFERENCE: 390069.401C1  
CURRENT APPLICATION NUMBER: US/10/207,655  
CURRENT FILING DATE: 2002-07-25  
NUMBER OF SEQ ID NOS: 426  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 139  
LENGTH: 739  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-207-655-139

Query Match 6.7%; Score 155; DB 14; Length 739;  
Best Local Similarity 21.9%; Pred. No. 0.0038;  
Matches 91; Conservative 71; Mismatches 166; Indels 88; Gaps 18;

QY RHLVLVQLALPPAT--QGNKVVLGKKGVLTCTASQKKSIOFHWKNSNOIKILGNQ 65  
DB 211 RQAVKEIQVIVSPKNTVISVNPSTKLOEGSVTWTCSSEGLPAPEIFW----- 258  
QY 66 GSFLTKGPSKLNDRASRLMDQGNFPLIKNLKIEDSDTYICE---VEDQKEEYQL 121  
DB 259 -----SKLIDNGNLQHL---SGNATLTLIAMRMEDSGIYVCEGVNLIGKRNKEVELI 307  
QY 122 V----FGLTANSDTHLLQ--GQSLTLTLSPSSPSVQCRSPRGKNIQG-----GKT 168

DB 308 VOEKPFTVEISPGPRIAQAIGDSVMLTCSVMGCESPFSFWRTQIDPSLGSKVSEGTNST 367  
QY 169 LSVSQAELQDSGWTCTVLAQNKVFEKIDIVLAFOKASSIYK---KEGEQVEFS--- 222  
DB 368 LTLSPVSEFENHSHYLCVTGCKHKKELQVELYSPRPDEIEMSGLVNGSSVTYVSCV 427  
QY 223 ---PFL-AFTVEKLTSGSELMWQERASSKSWITF---DLKKEVSVKRVTDPKLQ- 273  
DB 428 PSYVPLDRLEIELKGETIL-----ENIEFLIEDTDMKSLKMSLEMTPIPTIED 476  
QY 274 MGKLL-----PLHLTLPOALPOYAGSGNLTALBAKTKLHOEVNLVYMRATOLQK---N 325  
DB 477 TKGALVQOAKLHIDMEFEKQKQOS---TQTLVYVNAV--RDTVLVSPSSILEGSSVN 531  
QY 326 LTCSEWGPSTPKMLSLKENKAKVSKREKPVWVLPNPEAGMOCCLSDSGVLL 381  
DB 532 MTLCSQGFAPAKILMSRQLPNEGELQISENATLTLISTK-----MEDSGVLYCE 580

RESULT 54  
US-10-234-041-7  
Sequence 7, Application US/10234041  
Publication No. US20030153731A1  
GENERAL INFORMATION:  
APPLICANT: Biogen, Inc.  
APPLICANT: Hession, Catherine A.  
APPLICANT: Lobb, Roy R.  
APPLICANT: Goelz, Susan E.  
APPLICANT: Osborn, Laurelee  
APPLICANT: Benjamin, Christopher D.  
TITLE OF INVENTION: Endothelial Cell-Leukocyte Adhesion  
TITLE OF INVENTION: Molecules (ELAMs) and Molecules Involved in Leukocyte  
FILE REFERENCE: B124CP2DVZCN  
CURRENT APPLICATION NUMBER: US/10/234,041  
CURRENT FILING DATE: 2003-04-08  
PRIOR APPLICATION NUMBER: 08/473,764  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 08/342,642  
PRIOR FILING DATE: 1994-11-21  
PRIOR APPLICATION NUMBER: 07/608,298  
PRIOR FILING DATE: 1990-10-31  
PRIOR APPLICATION NUMBER: 07/452,675  
PRIOR FILING DATE: 1989-12-18  
PRIOR APPLICATION NUMBER: 07/359,516  
PRIOR FILING DATE: 1989-06-01  
PRIOR APPLICATION NUMBER: 07/345,151  
PRIOR FILING DATE: 1989-04-28  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 739  
TYPE: PRT  
ORGANISM: Homo sapien  
US-10-234-041-7

Query Match 6.7%; Score 155; DB 14; Length 739;  
Best Local Similarity 21.9%; Pred. No. 0.0038;  
Matches 91; Conservative 71; Mismatches 166; Indels 88; Gaps 18;

QY RHLVLVQLALPPAT--QGNKVVLGKKGVLTCTASQKKSIOFHWKNSNOIKILGNQ 65  
DB 211 RQAVKEIQVIVSPKNTVISVNPSTKLOEGSVTWTCSSEGLPAPEIFW----- 258  
QY 66 GSFLTKGPSKLNDRASRLMDQGNFPLIKNLKIEDSDTYICE---VEDQKEEYQL 121  
DB 259 -----SKLIDNGNLQHL---SGNATLTLIAMRMEDSGIYVCEGVNLIGKRNKEVELI 307  
QY 122 V----FGLTANSDTHLLQ--GQSLTLTLSPSSPSVQCRSPRGKNIQG-----GKT 168  
DB 308 VOEKPFTVEISPGPRIAQAIGDSVMLTCSVMGCESPFSFWRTQIDPSLGSKVSEGTNST 367



```

Oy      169  /SVSGLELQDSTWTCVLOQOKXVEFKIDIVLAFOKASIVYK---KEGQOVERS--- 222
Db      368  LTLSSVSENEHSYLCVTCGHKKLEKGIQVELYFPRDPELEMGGVLNSSVYVSKV 427
Oy      223  ---PPL-AFYVEKLTGSGELMMQARASSKSMITF---DLKNKEVSVKRVTDOPKLO- 273
Db      428  PSVYVLDLDELIELBLGFTLL-----ENIEFLBEDTDMKSLJENKSLEMTFPIPTED 476
Oy      274  MGKUL-----PLHLTLPQALPOYAGSGNLTLAELAKTGKHOEVLNVYMRATOLQK---N 325
Db      477  TGLALVCOAKLHIDIMEPEFKOROS---TOTLYVNVAP--RDTTVLVSSPSSILEGGSSVN 531
Oy      326  LTCEVWGPSPKMLMSLKLENKEAVSVKREKVWVNLNPEAGMOCCLSPSGQVLE 381
Db      532  MTLCSQGFPPAPKILMSRQLPNELOPLSENATLLTILSTK-----MEDSGYLCE 580

RESULT 55
US-10-231-956A-325
; Sequence 325, Application US/10231956A
; Publication No. US2004005323A1
; GENERAL INFORMATION:
; APPLICANT: Lorens, James B.
; APPLICANT: Xu, Weiduan
; APPLICANT: Bogenberger, Jakob
; APPLICANT: Holland, Sacha
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: Modulators of Angiogenesis
; FILE REFERENCE: 021044-004100US
; CURRENT APPLICATION NUMBER: US/10/231,956A
; CURRENT FILING DATE: 2001-08-30
; NUMBER OF SEQ ID NOS: 522
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 325
; LENGTH: 1479
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-231-956A-325

Query Match      6.5%; Score 152; DB 12; Length 1479;
Best Local Similarity 21.5%; Pred. No. 0.017;
Matches 100; Conservative 57; Mismatches 155; Indels 154; Gaps 19

Oy      34  GDTVELTCTASQSKSIQFHMKNSNQIKLNGSGFLTKGSPKSLNDRASRLMDQGNP 93
Db      260  GNTVYFTGRACENGPKEIILWNN-----NLSMKTDSRLMLDDGT-- 301
Oy      94  LIINKLAKIEDSDTYICEV-----EDQKEVQLLVEGLTAN-----SDTHLLOGSILT 141
Db      302  LMIQNTGCTDGOIYOCMAKNVAGVEKGTGTLRYGSPARPFVYIQPQNTVELVDESVTL 361
Oy      142  ---TLSEPPGSSPSVQCRSP---RGKNIQGGKTLVSQLELQDSTWTCVLOQOKV 193
Db      362  ECSAIGHPPRPRIKSWTRCDRTPLVPDPRVNIITPSGGVLIQNVVQDSGEVACATNNIDSV 421
Oy      194  EFKIDIVLAFOKASIVYK---EGEQVEFSPFLAFYVEKLTGSGELMMQARASSSKS 250
Db      422  HATAFIIQALRQPIFTVTPODRVVIEGQIVDF-----OEAKGNPP 462
Oy      251  WITPDLKNKEVSVK---VTQDPKLOM-----GKULPLHLTL--PO 286
Db      463  VIAMTKGGSQLSVDRRLHLVLSGTLRISGVALHDQGYEQCAVNIIGSKVVAHLTVQPR 522
Oy      287  ALPOYAGSGNLTLAELAKTGKHOEVLNVYMRATOLQKNLTGEVWGPTSPKMLMSLKEN 346
Db      523  VTPVPAASIPSDTV-----EVGANV-----QLPSSQGEPEPAITW-----N 559
Oy      347  KEAKSVKREKPVWVNLNPE-----AGMOCCL-----SPSGQVLESNIKVLP 389
Db      560  KDG-VQYVTEBSGCFHLSPEGFLLTINDVGPADAKRYECVARNITGSAVSVMVLSVN----- 613
Oy      390  WSTPHPRASALPAPPTGSALPDPOTASALPDPAPASALPALALAVI 435

```

```

Db          614 -----PD-----VSRRNDPFAVATISVEALATV   635

RESULT 56
US-10-211-462-87
; Sequence 87, Application US/10211462
; Publication NO. US20040033495A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynnne, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Aziz, Natasha
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Angiogenesis, Compositions and
; TITLE OF INVENTION: Methods of Screening for Angiogenesis Modulators
; FILE REFERENCE: 018501-006200US
; CURRENT APPLICATION NUMBER: US/10/211,462
; PRIOR FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: US 09/784,366
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/791,390
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 60/310,025
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/334,244
; PRIOR FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 87
; LENGTH: 1496
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-211-462-87

Query Match      6.5%; Score 152; DB 12; Length 1496;
Best Local Similarity 21.5%; Pred. No. 0.017; Indels 154; Gaps 19
Matches 100; Conservative 57; Mismatches 155;

QY    34 GDVVELTGTAAQKKSIOFMKNNSNOIKILGNQGSFLTGPSEKLDRADRSRLMDQGNFP 93
DB    277 GNTVYFTCRAGNRPKEIWLARN-----NELSMKDSRLNLDDGR-- 318
QY    94 LIINKLKIEDSDTYICEV---EDQKEVOLLVFGLTAN-----SDTHLLOGQSULTL 141
DB    319 LMIQNTGRTDGGIYCMAKVAVGEVKTGDEVTLYRFSGSPARPTFYIQPNTENVLGESVTL 378
QY    142 ----TLEPPSSPSVQCRRP---RGKNIGQGKTLSSQLDELDSGTWTCTVLONQKYV 193
DB    379 ECSATGHPRPISTWTRGDRTLPVPDPNRNITPSSGLYLQNVAODSGETYASATANNIDSV 438
QY    194 EPKLDIVVLARQKXASSIYYKK---EGEVEFPSFLAFVEKLTJSGELMMQAERASSSKS 250
DB    439 HATAFIYQALPQFVTWPQRDIVIEGQTVDF-----OCEAKGNPP 479
QY    251 WITEPLDKNKEYSVR---VTQDPKLOM-----GKCLPLHLTL-PQ 286
DB    480 VIAMTKGSGSLVDNRHVLVSSGTLRIISGVALLHDQGYEQCAVNIIISQKVVAAHLTVQPR 539
QY    287 ALPOYAGSGNLTALAEATGKTHOEVLNVNRATOLOKNTLTCBEWGWPTSPKLMISLKLLEN 346
DB    540 VTPVFASIPSDTTV-----EVGANV-----QLPCSSQGEPEPAITW----N 576
QY    347 KEAVYSKRKKVWVWLNPE-----AGMMQCLL-----SDSQGVLLSNINIKVLP 389
DB    577 KDG-VQVTESEKFHSPEGFLTINDVGPADRGREVCARRNTIGSASVMVLSLVNV----- 630
QY    390 WSTVHPRASALPARPTGSALPDQOTASALRDPDAASALPALALVI 435
DB    631 -----PD-----VSRRNDPFAVATISVEALATV 652

RESULT 57

```

```

US-10-021-660-125
; Sequence 125, Application US/10021660
; Publication No. US20030152926A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynn, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: EOS Biotechnology, Inc.
; TITLE OF INVENTION: No. US20030152926A1 Method of Diagnosis of Angiogenesis,
; TITLE OF INVENTION: Compositions and Methods of Screening for Angiogenesis,
; TITLE OF INVENTION: Modulators
; FILE REFERENCE: 018501-000710US
; CURRENT FILING DATE: US/10/021, 660
; PRIOR FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: US/09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 125
; LENGTH: 1496
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-021-660-125

```

Query Match 6.5%; Score 152; DB 14; Length 1496;

Best Local Similarity 21.5%; Pred. No. 0.017; Indels 154; Gaps 19;

Matches 100; Conservative 57; Mismatches 155;

```

QY 34 GDTVELTCTASQKSIQPHMKNSQIKILNGQSFLTQPSKLNDRADSRSLMDQGNFP 93
DB 277 GNTVYFTCRAGNPKPEIILWRNN-----NELSMKTDSRLNLLDDGT-- 318
QY 94 LIINKLKIEDSDTYICEV-----EDQKEVOLLVFGILTAN-----SDTHLLOGQSITL 141
DB 319 LMIQNTQETDQGIYQCMKKNVAGEVKTQETVLTAFYFGSPARPTFVIQPNTEVLVGSVTL 378
QY 142 -----TLSPGSSPSVQCRSP-----RGKNIQSGKTLVSQLELQDSGTWTCTVLOQNKV 193
DB 379 ECSATGHPRIISWTRGRTPLFVDPVRVNIITPSGGLYIQNVVQDSGEVACSAATNIDSV 438
QY 194 EFKIDIVVLAFOKASSIYVKK---EGEOVEFSPLAFTVEKLTGSGELMWAERASSKS 250
DB 439 HATAFIIVQALPQFTVTPQDRVIEGQTVDF-----QCEAKGNPPP 479
QY 251 WITFDLKNKEVSVK---VTQDPKLOM-----GKTLPLHLTL-PQ 286
DB 480 VIAWTKGSGSLSDVRHRLVLSGTLRISGVALHDQGYEQCAVNIIGSQKVAHLTVQPR 539
QY 287 ALPOYAGSGNLTLALAKTKGLHGEVNLVVMRATQLOKNLTCCEWGPPTSFKMLSLKLEN 346
DB 540 VTFVFASISPDITV-----EVGANV-----QLPCCSQGEPEPPIITW-----N 576
QY 347 KEAKVSKREKPVVNLNPE-----AGMWOCLL-----SDSGOVLLESNIKVLPT 389
DB 577 KDG-VQVYTESGKPHISPEGFLTINDVGPADAGRECVARNTIGSASVSVLNVN----- 630
QY 390 WSTFVHPRASALPAPPTGSALPDPTASALPDPPASALPALAVI 435
DB 631 -----PD-----VSRNGDPFVATSIYEAIATV 652

```

RESULT 58

US-10-331-496A-28

; Sequence 28, Application US/10331496A

; Publication No. US2003028305A1

; GENERAL INFORMATION:

; APPLICANT: FRANTZ, GRETCHEN

; APPLICANT: HILLAN, KENNETH J.

; APPLICANT: PHILLIPS, HEIDI S.

; APPLICANT: POLAKIS, PAUL

; APPLICANT: SMITH, VICTORIA

```

; APPLICANT: SPENCER, SUSAN D.
; APPLICANT: WILLIAMS, P. MICKEY
; APPLICANT: WU, THOMAS D.
; APPLICANT: ZHANG, ZEMIN
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF TUMOR
; FILE REFERENCE: P5014R1-PCT
; CURRENT APPLICATION NUMBER: US/10/331,496A
; CURRENT FILING DATE: 2002-12-30
; PRIOR APPLICATION NUMBER: US 60/345,444
; PRIOR FILING DATE: 2002-01-02
; PRIOR APPLICATION NUMBER: US 60/351,885
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: US 60/360,066
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: US 60/362,004
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/366,869
; PRIOR FILING DATE: 2002-03-20
; PRIOR APPLICATION NUMBER: US 60/366,284
; PRIOR FILING DATE: 2002-03-21
; PRIOR APPLICATION NUMBER: US 60/368,679
; PRIOR FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: US 60/404,809
; PRIOR FILING DATE: 2002-08-19
; PRIOR APPLICATION NUMBER: US 60/405,645
; PRIOR FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 95
; SEQ ID NO 28
; LENGTH: 1496
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-331-496A-28

```

Query Match 6.5%; Score 152; DB 15; Length 1496;

Best Local Similarity 21.5%; Pred. No. 0.017; Indels 154; Gaps 19;

Matches 100; Conservative 57; Mismatches 155;

```

QY 34 GDTVELTCTASQKSIQPHMKNSQIKILNGQSFLTQPSKLNDRADSRSLMDQGNFP 93
DB 277 GNTVYFTCRAGNPKPEIILWRNN-----NELSMKTDSRLNLLDDGT-- 318
QY 94 LIINKLKIEDSDTYICEV-----EDQKEVOLLVFGILTAN-----SDTHLLOGQSITL 141
DB 319 LMIQNTQETDQGIYQCMKKNVAGEVKTQETVLTAFYFGSPARPTFVIQPNTEVLVGSVTL 378
QY 142 -----TLSPGSSPSVQCRSP-----RGKNIQSGKTLVSQLELQDSGTWTCTVLOQNKV 193
DB 379 ECSATGHPRIISWTRGRTPLFVDPVRVNIITPSGGLYIQNVVQDSGEVACSAATNIDSV 438
QY 194 EFKIDIVVLAFOKASSIYVKK---EGEOVEFSPLAFTVEKLTGSGELMWAERASSKS 250
DB 439 HATAFIIVQALPQFTVTPQDRVIEGQTVDF-----QCEAKGNPPP 479
QY 251 WITFDLKNKEVSVK---VTQDPKLOM-----GKTLPLHLTL-PQ 286
DB 480 VIAWTKGSGSLSDVRHRLVLSGTLRISGVALHDQGYEQCAVNIIGSQKVAHLTVQPR 539
QY 287 ALPOYAGSGNLTLALAKTKGLHGEVNLVVMRATQLOKNLTCCEWGPPTSFKMLSLKLEN 346
DB 540 VTFVFASISPDITV-----EVGANV-----QLPCCSQGEPEPPIITW-----N 576
QY 347 KEAKVSKREKPVVNLNPE-----AGMWOCLL-----SDSGOVLLESNIKVLPT 389
DB 577 KDG-VQVYTESGKPHISPEGFLTINDVGPADAGRECVARNTIGSASVSVLNVN----- 630
QY 390 WSTFVHPRASALPAPPTGSALPDPTASALPDPPASALPALAVI 435
DB 631 -----PD-----VSRNGDPFVATSIYEAIATV 652

```

RESULT 59

US-10-276-774-1957

```
; Sequence 1957, Application US/10276774
; Publication No. US20040053245A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; APPLICANT: Tang, Y. Tom et al
; TITLE OF INVENTION: No. US20040053245A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 21272-030
; CURRENT APPLICATION NUMBER: US/10/276,774
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 2700
; SOFTWARE: Custom
; SEQ ID NO 1957
; LENGTH: 1498
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-276-774-1957

Query Match      6.5%; Score 152; DB 12; Length 1498;
Best Local Similarity 21.5%; Pred. No. 0.017;
Matches 100; Conservative 57; Mismatches 155; Indels 154; Gaps 19;

QY 34 GDTVELTCTASQKSIQFHWKNSNOIKLGNQGSFLTKGPKSLNDRADSRSLMDQGNP 93
DB 277 GNTVFTCTCAEGNPKREIIMLRNN-----NELSMKTD SRLNLDDGT-- 318
QY 94 LIINKLKIEDSDTYICEV-----EDQKEVQLLVFGLTAN-----SDTHLLQGQSITL 141
DB 319 LMIQNTQETDQGIYQCAKNVAGVEKTVLRYFGSPARPTFYIQPONTVEVLVGESVTL 378
QY 142 ----TLESPPSSPSVQCRSP----RGKNIQSGKTLVSQLELDQSGTWCTTVLONQKV 193
DB 379 ECSATGHPRPRIISWTRGRTPLPYDPRVNITPSGGLYIQNVQDSEGYACSAATNNIDSV 438
QY 194 EFKIDIVVLAFOKASISIVKK---EGEQVEFSPLAFTVEKLTGSGELMWQAEASSSKS 250
DB 439 HATAFIIYQALPOFTVTRQDRVIEGQTVDF-----QCEAKGNPP 479
QY 251 WITFDLKNKEVSVKR---VTODPKLQM-----GKULPLHLTL-PQ 286
DB 480 VIAMTKGSGQLSVDRHLVLSGTLRISGVALHDQGYECQAVNIIGSKVVAHLTVQPR 539
QY 287 ALPOYAGSGNLTLEAKTGKLGHEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLEN 346
DB 540 VTPVFASIPSDTTV-----EYGANV-----QLPSSQGEPEBPITW-----N 576
QY 347 KEAKVSKREKPVWVNLPE-----AGMOCCL-----SDSGQVLESNIKVLPT 389
DB 577 KQG-VQVTESGKFHISPEGFLTINDVGPADAGRYECVARNITIGSASVSMVLNVN----- 630
QY 390 WSTPVHPRASALPAPPTGSALPDPTGASALPDPPAASALPALAVI 435
DB 631 -----PD---VSRNGDPVATSIYEALATV 652

RESULT 60
US-10-243-552-899
; Sequence 899, Application US/10243552
; Publication No. US20030224379A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Meng, Gezhi
; APPLICANT: Ma, Yungang
; TITLE OF INVENTION: Novel Nucleic Acids and
; FILE REFERENCE: 807A
; CURRENT APPLICATION NUMBER: US/10/243,552
; CURRENT FILING DATE: 2002-09-12
```

```
; PRIOR APPLICATION NUMBER: US 60/322,511
; PRIOR FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/04927
; PRIOR FILING DATE: 2001-02-26
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 998
; SOFTWARE: pc_fl_genes Version 5.0
; SEQ ID NO 899
; LENGTH: 1498
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-243-552-899

Query Match      6.5%; Score 152; DB 12; Length 1498;
Best Local Similarity 21.5%; Pred. No. 0.017;
Matches 100; Conservative 57; Mismatches 155; Indels 154; Gaps 19;

QY 34 GDTVELTCTASQKSIQFHWKNSNOIKLGNQGSFLTKGPKSLNDRADSRSLMDQGNP 93
DB 277 GNTVFTCTCAEGNPKREIIMLRNN-----NELSMKTD SRLNLDDGT-- 318
QY 94 LIINKLKIEDSDTYICEV-----EDQKEVQLLVFGLTAN-----SDTHLLQGQSITL 141
DB 319 LMIQNTQETDQGIYQCAKNVAGVEKTVLRYFGSPARPTFYIQPONTVEVLVGESVTL 378
QY 142 ----TLESPPSSPSVQCRSP----RGKNIQSGKTLVSQLELDQSGTWCTTVLONQKV 193
DB 379 ECSATGHPRPRIISWTRGRTPLPYDPRVNITPSGGLYIQNVQDSEGYACSAATNNIDSV 438
QY 194 EFKIDIVVLAFOKASISIVKK---EGEQVEFSPLAFTVEKLTGSGELMWQAEASSSKS 250
DB 439 HATAFIIYQALPOFTVTRQDRVIEGQTVDF-----QCEAKGNPP 479
QY 251 WITFDLKNKEVSVKR---VTODPKLQM-----GKULPLHLTL-PQ 286
DB 480 VIAMTKGSGQLSVDRHLVLSGTLRISGVALHDQGYECQAVNIIGSKVVAHLTVQPR 539
QY 287 ALPOYAGSGNLTLEAKTGKLGHEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLEN 346
DB 540 VTPVFASIPSDTTV-----EYGANV-----QLPSSQGEPEBPITW-----N 576
QY 347 KEAKVSKREKPVWVNLPE-----AGMOCCL-----SDSGQVLESNIKVLPT 389
DB 577 KQG-VQVTESGKFHISPEGFLTINDVGPADAGRYECVARNITIGSASVSMVLNVN----- 630
QY 390 WSTPVHPRASALPAPPTGSALPDPTGASALPDPPAASALPALAVI 435
DB 631 -----PD---VSRNGDPVATSIYEALATV 652

RESULT 61
US-10-076-674-4
; Sequence 4, Application US/10076674
; Publication No. US20030165478A1
; GENERAL INFORMATION:
; APPLICANT: Sokoll, Kenneth K.
; TITLE OF INVENTION: Stabilized Synthetic Immunogen Delivery System
```

```

; FILE REFERENCE: Immunogen Delivery System
; CURRENT APPLICATION NUMBER: US/10/076,674
; CURRENT FILING DATE: 2002-04-23
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (20)..(20)
; OTHER INFORMATION: Xaa indicates epsilon-Lys
US-10-076-674-4

Query Match
Best Local Similarity 100.0%; Score 150; DB 14; Length 50;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 NQSFLLTKGPKSLNDRADSRSLMDQGN 91
Db 22 NQSFLLTKGPKSLNDRADSRSLMDQGN 49

RESULT 62
US-10-076-674-5
; Sequence 5, Application US/10076674
; Publication No. US20030165478A1
; GENERAL INFORMATION:
; APPLICANT: Sokoll, Kenneth K.
; TITLE OF INVENTION: Stabilized Synthetic Immunogen Delivery System
; FILE REFERENCE: Immunogen Delivery System
; CURRENT APPLICATION NUMBER: US/10/076,674
; CURRENT FILING DATE: 2002-04-23
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (20)..(20)
; OTHER INFORMATION: Xaa indicates epsilon-Lys
US-10-076-674-5

Query Match
Best Local Similarity 100.0%; Score 150; DB 14; Length 50;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 NQSFLLTKGPKSLNDRADSRSLMDQGN 91
Db 22 NQSFLLTKGPKSLNDRADSRSLMDQGN 49

RESULT 63
US-10-355-161A-4
; Sequence 4, Application US/10355161A
; Publication No. US20040009897A1
; GENERAL INFORMATION:
; APPLICANT: Sokoll, Kenneth K.
; TITLE OF INVENTION: Stabilized Synthetic Immunogen Delivery System
; FILE REFERENCE: Immunogen Delivery System
; CURRENT APPLICATION NUMBER: US/10/355,161A
; CURRENT FILING DATE: 2003-01-31
; PRIOR APPLICATION NUMBER: US 10/076674
; PRIOR FILING DATE: 2002-02-14
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Human
```

```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (20)..(20)
; OTHER INFORMATION: Xaa indicates epsilon-Lys
US-10-355-161A-4

Query Match
Best Local Similarity 100.0%; Score 150; DB 15; Length 50;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 NQSFLLTKGPKSLNDRADSRSLMDQGN 91
Db 22 NQSFLLTKGPKSLNDRADSRSLMDQGN 49

RESULT 64
US-10-355-161A-5
; Sequence 5, Application US/10355161A
; Publication No. US20040009897A1
; GENERAL INFORMATION:
; APPLICANT: Sokoll, Kenneth K.
; TITLE OF INVENTION: Stabilized Synthetic Immunogen Delivery System
; FILE REFERENCE: Immunogen Delivery System
; CURRENT APPLICATION NUMBER: US/10/355,161A
; CURRENT FILING DATE: 2003-01-31
; PRIOR APPLICATION NUMBER: US 10/076674
; PRIOR FILING DATE: 2002-02-14
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (20)..(20)
; OTHER INFORMATION: Xaa indicates epsilon-Lys
US-10-355-161A-5

Query Match
Best Local Similarity 100.0%; Score 150; DB 15; Length 50;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 64 NQSFLLTKGPKSLNDRADSRSLMDQGN 91
Db 22 NQSFLLTKGPKSLNDRADSRSLMDQGN 49

RESULT 65
US-10-274-583-20
; Sequence 20, Application US/10274583
; Publication No. US20030138431A1
; GENERAL INFORMATION:
; APPLICANT: Exelixis, Inc.
; TITLE OF INVENTION: LRCAPs AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-119C
; CURRENT APPLICATION NUMBER: US/10/274,583
; CURRENT FILING DATE: 2002-10-21
; PRIOR APPLICATION NUMBER: 60/338,733
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/357,600
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 60/361,196
; PRIOR FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 1477
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-274-583-20

Query Match
Best Local Similarity 100.0%; Score 146; DB 14; Length 1477;
```

Best Local Similarity 22.2%; Pred. No. 0.049;  
Matches 88; Conservative 49; Mismatches 140; Indels 120; Gaps 16;

QY 34 GGTVELTCTASQKKSIOFHMKNSNQIKLGNOSFLTGPBKLNDPDRSRSLMDQGNP 93  
DB 260 GATVFTCTCAEGNPKBEIIMLRNN-----NELSMKTDRLMLDDGT-- 301  
QY 94 LIHKNLKIDSPTIYCEV-----EDQKEEYOLLVFGLTAN-----SDTHLLOGSLTL 141  
DB 302 LHIQNTQETDQGIYQCMACKNAGEVKTQEVTLTAFSSPAPPTFVIOPTQTEVLVGEVTL 361  
QY 142 ----TLESPPGSSPSVQCRSP-----RGKNIQGGKTLVSQLELQDSGTCTVLQONQKY 193  
DB 362 ECSATGHPPEPRISWTRGDRPLPVDRPVNITPSGGLYQNVQGSBGEVACSATNNIDV 421  
QY 194 ERKIDVLVAFQKASSIVYK--EEOYEFSEFLPAFTVEKLTGSGELMWQERASSSKS 250  
DB 422 HATTAFTIYQALPQFTYTPQDRVVEGQTVDF-----QCEAKGNPP 462  
QY 251 WITFDLKNKEVSVK---VTQDPKLOM-----GKKLPLHLTL-PQ 286  
DB 463 VLAWTGSGQLSVDRHVLVSSGTLKISGVALHDQGCQECQAVNITGSKVVAHLTVQPR 522  
QY 287 ALPQVAGSNLTALAEAKTGKLEHVEVNLVVRATQLOKNLTCEVWGPTSPKMLSLKEN 346  
DB 523 VTPVFASIPSDTV-----EVGANV-----QLPSSQGEPPPATW-----N 559  
QY 347 KEAKVSKREKPVVNLNPE-----AGMMOCL 371  
DB 560 KQG-VQTESGKFHISPEGFLTINDVGPADAGRYECV 595

RESULT 66  
US-10-168-417A-4  
; Sequence 4, Application US/10168417A  
; Publication No. US20040009185A1  
; GENERAL INFORMATION:  
; APPLICANT: Emberg, Peter  
; APPLICANT: Barber, Brian  
; APPLICANT: Sambhara, Suryprakash  
; APPLICANT: Sia, Charles Duo Yuan  
; TITLE OF INVENTION: Enhancing the Immune Response to an Antigen by Presensitizing with  
; TITLE OF INVENTION: Inducing Agent Prior to Immunizing with the Inducing Agent and  
; FILE REFERENCE: 11014-18-US  
; CURRENT APPLICATION NUMBER: US/10168,417A  
; CURRENT FILING DATE: 2002-11-15  
; PRIOR APPLICATION NUMBER: US 60/174,587  
; PRIOR FILING DATE: 2000-01-05  
; PRIOR APPLICATION NUMBER: PCT/CA01/00005  
; PRIOR FILING DATE: 2001-01-05  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 4  
; LENGTH: 701  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: modified CEA  
US-10-168-417A-4

Query Match 6.2%; Score 145; DB 15; Length 701;  
Best Local Similarity 18.9%; Pred. No. 0.021;  
Matches 96; Conservative 75; Mismatches 216; Indels 122; Gaps 19;

DB 246 NTSVSGENMLNSCHA--ASNPPAQSWFVANGTPOQSOQLFIPRITVNNSGSYTCQAHN 303  
QY 189 NOKKVEFKIDIVLAFQKASSIVYKKEGOYEFSEFLPAFTVE-KUTSGELMWQERASS 247  
DB 304 SDTGINRTVTVTITVYEEPPKFPITSNNSNPVEDEDAVALTCEPTEIONTTTLMWVNNQSLP 363  
QY 248 SKSWITFDLKNKEVSVKRVTDQP-----KIQMKKPLHL-----TLPQAL 288  
DB 364 VSPRLQSLNDKNTLTLVTRNDVGPYECGIQNELSVHSDPVLINVLVGPDDPTISPSY 423  
QY 289 PQYAGSNLTALAEAKT-----GKLEHVE-NLVMRATQLOKNLTCEVWGPTSP 335  
DB 424 TYRRGVNLISLSCHAASNPAPQYSLWLDIGNIQHTQELFIGNITEKNSGLVTCQANNAS 483  
QY 336 PKMLSLKLENKAKV-----SKREKV-----WVLNPPAGMMOCLSDSGVLLS 382  
DB 484 GHSRTTVKTIIVSALPKPSISSNNSKVEDKDAVAFCEBEAQVTVLMMVNNQ----- 538  
QY 383 NIKVLPWSTPRHPPASALPAPT-----GSALDPQTRASAL- 419  
DB 539 -----SLVSPRLQSLNNGNRITLTFNTRNDARAYVCGIQNSVSNRSDPVTLDVLY 590  
QY 420 -PDPPAASALPALAVISFLGLGIVAC 447  
DB 591 GPDPTIISPPDS-----SYLSGADLNLSC 614

RESULT 67  
US-10-176-847-32  
; Sequence 32, Application US/10176847  
; Publication No. US20030068636A1  
; GENERAL INFORMATION:  
; APPLICANT: Veiby, Peter Ole  
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR  
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST  
; FILE REFERENCE: MRI-039  
; CURRENT APPLICATION NUMBER: US/10/176,847  
; CURRENT FILING DATE: 2002-06-21  
; NUMBER OF SEQ ID NOS: 112  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 32  
; LENGTH: 647  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-176-847-32

Query Match 6.2%; Score 143; DB 14; Length 647;  
Best Local Similarity 20.9%; Pred. No. 0.027;  
Matches 105; Conservative 81; Mismatches 151; Indels 166; Gaps 25;

QY 16 LALLPAAQGNKV-----VLGKGDVTELTCTASQKKSIOFHMKNSNQIKLGN--NQ 65  
DB 15 LMIWMAASQAFKIETTPESRYLAQIGDSVSLTCTGCGESPFPSKRQIDPSLNGKVTNE 74  
QY 66 G-SFLTGPBKLNDP-----ADSRSLMDQGS-----NFP----- 93  
DB 75 GTTSTLTWNPVSPFGNHSYLCTATCESRK--LEKGIQVEIYSPDPDEPHLSGLEAKGP 132  
QY 94 -----LIHKNLK-----IIDS-----TYICEVED----- 113  
DB 133 ITVKGSVADVPPRRLIEDLKGHLMSQFLEBDARKSLKTSLEVTFPVEDIGKV 192  
QY 114 -----OKEEVOLLVFG-----LTPNSDTHLLOGSLTLTLESPPGS 149  
DB 193 LVCRAKLHIDEMDSVPTVQAVKELQYISKNTVVISNPSKLOEGBSVMTCCSSELP 252  
QY 150 SPVSV---QCRSPGKNIQGGKTLVSQLELQDSGTWC---TVLQONKKEFKIDIV 201  
DB 253 APEIFWSKLDNGNLIQHLSGNATVTLTAMRVEDSGIYCEGVNLIGKRRKEV---LIV 308  
QY 202 LAFQKASSIVYK--EEOYEFSEFLPAFTVEKLTGSGELMWQERASSSKS 251

```
Db 309 QAFPRDEIMSGLVNGSSVTWSCVPSVYPLDRLEIELKGETIL-----EN 357
Qy 252 ITF-----DLKNKEVSKRVYQDPKLO-MGKUL-----PLHLTPQALPOVAGSGLTTLALE 302
Db 358 IEFLEDTMKSLEKSLKEMTFPTIEDTGKALVCQAKLHIDMEFEPPKORS---TOTLY 414
Qy 303 AKTGKLEHVEVNLVVRATOLQK---NLTCVWGPTSPKMLSLKLENKEAKVSKREKPV 358
Db 415 VNVAP--RDTTVLVSPSSILEBSSVNMTCISQGFAPAKILMSRQLPNEGLOPLSENATL 472
Qy 359 WLVNPEAGMWQCLISDSGOVLE 381
Db 473 TLISTK-----MEDSGVYLCE 488
```

```
RESULT 68
US-10-234-041-6
/ Sequence 6, Application US/10234041
/ Publication No. US20030153731A1
/ GENERAL INFORMATION:
/ APPLICANT: Biogen, Inc.
/ APPLICANT: Heesion, Catherine A.
/ APPLICANT: Lobb, Roy R.
/ APPLICANT: Goelz, Susan E.
/ APPLICANT: Osborn, Laurelee
/ APPLICANT: Benjamin, Christopher D.
/ APPLICANT: Rosa, Margaret D.
/ TITLE OF INVENTION: Endothelial Cell-Leukocyte Adhesion
/ TITLE OF INVENTION: Molecules (ELAMs) and Molecules Involved in Leukocyte
/ FILE REFERENCE: B124CP2DV2CN
/ CURRENT FILING DATE: 2003-04-08
/ PRIOR APPLICATION NUMBER: US/10/234,041
/ PRIOR FILING DATE: 1995-06-07
/ PRIOR APPLICATION NUMBER: 08/473,764
/ PRIOR FILING DATE: 1995-06-07
/ PRIOR APPLICATION NUMBER: 08/342,642
/ PRIOR FILING DATE: 1994-11-21
/ PRIOR APPLICATION NUMBER: 07/608,298
/ PRIOR FILING DATE: 1990-10-31
/ PRIOR APPLICATION NUMBER: 07/452,675
/ PRIOR FILING DATE: 1989-12-18
/ PRIOR APPLICATION NUMBER: 07/359,516
/ PRIOR FILING DATE: 1989-06-01
/ PRIOR APPLICATION NUMBER: 07/345,151
/ PRIOR FILING DATE: 1989-04-28
/ NUMBER OF SEQ ID NOS: 28
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 6
/ LENGTH: 647
/ TYPE: PRT
/ ORGANISM: Homo Sapien
US-10-234-041-6
```

```
Query Match 6.2%; Score 143; DB 14; Length 647;
Best Local Similarity 20.9%; Pred. No. 0.027;
Matches 105; Conservative 81; Mismatches 151; Indels 166; Gaps 25;

Qy 16 LALPAAQGNKV-----VLGKGDVVELTCTASQKKSIOFHKNKNQKIIG--NQ 65
Db 15 LWIMPAASQAFIETPTPSRYLAQIGDSVSLTCTGCESPFFSWRQIDSPLNKQVTNE 74
Qy 66 G--SFLTGKPSKLANDR-----ADSRSLMDQ-----NFP----- 93
Db 75 GTTSLTNPNVPSFGHEHSYLCATNCESRK--LEKIQVEIYSFPKDPRIHLSGPLEAKP 132
Qy 94 -----LIINKLK-----IEDSD-----TYICEVED----- 113
Db 133 ITVKCSADVVPFDRLEIDLKGDHLMKSGFLEADARKSLKTSLEVTFTFVIEDIGKV 192
Qy 114 -----OKEVQLLVFG-----LTNSDTHLQGSLLTLLESPPGS 149
Db 193 LVCRAKLHIDEMSVTVRQAVKELQVYISPKNTVIVSNPSTKLOEGGSSVTMTCSSEGLP 252
```

```
Qy 150 SPEV-----QCRSPRGKNIQCKTLVSQLELODSTWTG---TVLQNKCKVEFKIDIV 201
Db 253 APFLFMSKLDNNGNLQHSNATLTLTAMMBESGIVYCEGVNLIGNRREVE-----LIV 308
Qy 202 LAFQAKSIYK--KEGEVFS-----FPL-AFTVEKLTSSGELMWOAERASSKSW 251
Db 309 QAFPRDEIMSGLVNGSSVTWSCVPSVYPLDRLEIELKGETIL-----EN 357
Qy 252 ITF-----DLKNKEVSKRVYQDPKLO-MGKUL-----PLHLTPQALPOVAGSGLTTLALE 302
Db 358 IEFLEDTMKSLEKSLKEMTFPTIEDTGKALVCQAKLHIDMEFEPPKORS---TOTLY 414
Qy 303 AKTGKLEHVEVNLVVRATOLQK---NLTCVWGPTSPKMLSLKLENKEAKVSKREKPV 358
Db 415 VNVAP--RDTTVLVSPSSILEBSSVNMTCISQGFAPAKILMSRQLPNEGLOPLSENATL 472
Qy 359 WLVNPEAGMWQCLISDSGOVLE 381
Db 473 TLISTK-----MEDSGVYLCE 488
```

```
RESULT 69
US-10-032-189-128
/ Sequence 128, Application US/10032189
/ Publication No. US20030170630A1
/ GENERAL INFORMATION:
/ APPLICANT: Alabrook II, John P
/ APPLICANT: Tchernev, Velizar T
/ APPLICANT: Liu, Xiaohong
/ APPLICANT: Spytek, Kimberly A
/ APPLICANT: Zerhusen, Bryan D
/ APPLICANT: Patnrajan, Meera
/ APPLICANT: Grose, William M
/ APPLICANT: Lepley, Denise M
/ APPLICANT: Burgess, Catherine E
/ APPLICANT: Shinkets, Richard A
/ APPLICANT: Grose, William M
/ APPLICANT: Szekeres, Edward S
/ APPLICANT: Verneet, Corine A.M.
/ APPLICANT: Li, Li
/ APPLICANT: Caeman, Stacie J
/ APPLICANT: Boldog, Ferenc L
/ APPLICANT: Gorman, Linda
/ APPLICANT: Gangolli, Raha A
/ APPLICANT: Fernandes, Elma R
/ APPLICANT: Rieger, Daniel K
/ APPLICANT: Edinger, Shiomit R
/ APPLICANT: Gunther, Erik
/ APPLICANT: Miller, Isabelle
/ APPLICANT: Sciore, Paul
/ APPLICANT: Ellerman, Karen
/ APPLICANT: MacDougall, John R
/ APPLICANT: Smilison, Glenda
/ TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
/ FILE REFERENCE: 21402-228
/ CURRENT APPLICATION NUMBER: US/10/032,189
/ CURRENT FILING DATE: 2001-12-21
/ PRIOR APPLICATION NUMBER: 60/257,495
/ PRIOR FILING DATE: 2000-12-21
/ PRIOR APPLICATION NUMBER: 60/258,171
/ PRIOR FILING DATE: 2000-12-20
/ PRIOR APPLICATION NUMBER: 60/269,940
/ PRIOR FILING DATE: 2001-02-20
/ PRIOR APPLICATION NUMBER: 60/274,192
/ PRIOR FILING DATE: 2001-03-08
/ PRIOR APPLICATION NUMBER: 60/277,826
/ PRIOR FILING DATE: 2001-03-22
/ PRIOR APPLICATION NUMBER: 60/279,840
/ PRIOR FILING DATE: 2001-03-29
/ PRIOR APPLICATION NUMBER: 60/282,981
/ PRIOR FILING DATE: 2001-04-11
/ PRIOR APPLICATION NUMBER: 60/283,656
```

```
/ PRIOR FILING DATE: 2001-04-13
/ PRIOR APPLICATION NUMBER: 60/309,247
/ PRIOR FILING DATE: 2001-07-31
/ PRIOR APPLICATION NUMBER: 60/311,754
/ PRIOR FILING DATE: 2001-08-17
/ PRIOR APPLICATION NUMBER: 60/313,331
/ PRIOR FILING DATE: 2001-08-17
/ NUMBER OF SEQ ID NOS: 260
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO: 128
/ LENGTH: 5636
/ TYPE: PR
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: VARIANT
/ LOCATION: (3003)
/ OTHER INFORMATION: wherein Xaa is any amino acid.
/ FEATURE:
/ NAME/KEY: VARIANT
/ LOCATION: (3041)
/ OTHER INFORMATION: wherein Xaa is any amino acid.
/ FEATURE:
/ NAME/KEY: VARIANT
/ LOCATION: (3367)
/ OTHER INFORMATION: wherein Xaa is any amino acid.
US-10-032-189-128
```

```
Query Match 6.2%; Score 143; DB 14; Length 5636;
Best Local Similarity 20.4%; Pred. No. 0.53;
Matches 93; Conservative 56; Mismatches 168; Indels 138; Gaps 21;
```

```
QY 34 GDTVELTCTAASQKSIQFHMKNKSNQIKILGNQGSFLTPKPSKLNDRADSRSLMDQGNFP 93
DB 807 GSNVTLPCYVGQYEPPTIKMRRLDMPIFSR--PFSVSSISQLRTGA----- 851
QY 94 LIINKLKIEDSDTYICEVEDQ---KEEVQLLVFGLTA-----NSDTHLQGSGLTTL 143
DB 852 LFIILNLMASDKGTGYICEAENQFGKIQSEFTVTYGLVAPLIGISPSVANVIEGQQLTIPC 911
QY 144 ESPPGS-----SPSVQCRSPRGKNIQSGKTLVSQLELDQSGTWTG--- 184
DB 912 TLLAGNPPIPERRMWIKNSAMLLQNPITYRS-----DGSLLIERVOLQDGGEYTCVAS 963
QY 185 TYLQNKQKVEFKIDIVVLAFOKASSIVYKKEGEQVEFSPLAFYVEKLTGSGELMWQAR 244
DB 964 NAGTNNKTTSVVHVLPFIQHQQLLSTIEG--IPVTLP-----CKA 1004
QY 245 ASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHTLPQALPYAGSGNLTLALEAK 304
DB 1005 SGNPKPSVYWSKKGELIS---TSSAKFSAGADGSLVYVSPBG---EESGEYVCTATNT 1056
QY 305 TGKLNQEVNLVV-----MRATQLQK-----NLTCGV-----W--- 331
DB 1057 AGVAKRKQVLTYYVPRVFGDLRGLSDQPRVEISVLAGEEVTLLPCEVKSLEPPIITWAKE 1116
QY 332 ----GPTSK--LMSLKLENKEAKYSKKEKRWVVLNPPAGMWOCILSD-SCQV--LLES 382
DB 1117 TGLISPFSPRHFTLPBGSMKKTITRTS-----DSGYLVCATNIAGNVTAQAVL 1165
QY 383 NIKVLPTWST-PVHPRASA-----LPAPPTGSALP 411
DB 1166 NVHVPKIQGRPKHLKVQGVQVDFPCNAQGTPLP 1200
```

```
RESULT 70
US-10-120-801-72
/ Sequence 72, Application US/10120801
/ Publication No. US20030203843A1
/ GENERAL INFORMATION:
/ APPLICANT: Pena, Carol
/ APPLICANT: Guo, Xiaojia
/ APPLICANT: Shimkete, Richard
/ APPLICANT: Padigaru, Muralidhara
```

```
/ APPLICANT: Kekuda, Ramesh
/ APPLICANT: Spyrek, Kimberly
/ APPLICANT: Mehraban, Fuad
/ APPLICANT: Topper, James N.
/ APPLICANT: Malyankar, Uriel
/ APPLICANT: Wasserman, Scott
/ APPLICANT: Edinger, Shlomit
/ APPLICANT: Smithson, Glenda
/ APPLICANT: Gunther, Erik
/ APPLICANT: Komuves, Laszlo
/ TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
/ FILE REFERENCE: 21402-340
/ CURRENT APPLICATION NUMBER: US/10/120,801
/ PRIOR FILING DATE: 2002-04-11
/ PRIOR APPLICATION NUMBER: 60/285748
/ PRIOR FILING DATE: 2001-04-23
/ PRIOR APPLICATION NUMBER: 60/286068
/ PRIOR FILING DATE: 2001-04-24
/ PRIOR APPLICATION NUMBER: 60/286292
/ PRIOR FILING DATE: 2001-04-25
/ PRIOR APPLICATION NUMBER: 60/288334
/ PRIOR FILING DATE: 2001-05-03
/ PRIOR APPLICATION NUMBER: 60/291241
/ PRIOR FILING DATE: 2001-05-16
/ PRIOR APPLICATION NUMBER: 60/322284
/ PRIOR FILING DATE: 2001-09-14
/ PRIOR APPLICATION NUMBER: 60/285609
/ PRIOR FILING DATE: 2001-04-20
/ NUMBER OF SEQ ID NOS: 155
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO: 72
/ LENGTH: 5636
/ TYPE: PR
/ ORGANISM: human
/ FEATURE:
/ NAME/KEY: VARIANT
/ LOCATION: (3003)
/ OTHER INFORMATION: wherein Xaa is any amino acid.
/ FEATURE:
/ NAME/KEY: VARIANT
/ LOCATION: (3041)
/ OTHER INFORMATION: wherein Xaa is any amino acid.
/ FEATURE:
/ NAME/KEY: VARIANT
/ LOCATION: (3367)
/ OTHER INFORMATION: wherein Xaa is any amino acid.
US-10-120-801-72
```

```
Query Match 6.2%; Score 143; DB 15; Length 5636;
Best Local Similarity 20.4%; Pred. No. 0.53;
Matches 93; Conservative 56; Mismatches 168; Indels 138; Gaps 21;
```

```
QY 34 GDTVELTCTAASQKSIQFHMKNKSNQIKILGNQGSFLTPKPSKLNDRADSRSLMDQGNFP 93
DB 807 GSNVTLPCYVGQYEPPTIKMRRLDMPIFSR--PFSVSSISQLRTGA----- 851
QY 94 LIINKLKIEDSDTYICEVEDQ---KEEVQLLVFGLTA-----NSDTHLQGSGLTTL 143
DB 852 LFIILNLMASDKGTGYICEAENQFGKIQSEFTVTYGLVAPLIGISPSVANVIEGQQLTIPC 911
QY 144 ESPPGS-----SPSVQCRSPRGKNIQSGKTLVSQLELDQSGTWTG--- 184
DB 912 TLLAGNPPIPERRMWIKNSAMLLQNPITYRS-----DGSLLIERVOLQDGGEYTCVAS 963
QY 185 TYLQNKQKVEFKIDIVVLAFOKASSIVYKKEGEQVEFSPLAFYVEKLTGSGELMWQAR 244
DB 964 NAGTNNKTTSVVHVLPFIQHQQLLSTIEG--IPVTLP-----CKA 1004
QY 245 ASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHTLPQALPYAGSGNLTLALEAK 304
DB 1005 SGNPKPSVYWSKKGELIS---TSSAKFSAGADGSLVYVSPBG---EESGEYVCTATNT 1056
QY 305 TGKLNQEVNLVV-----MRATQLQK-----NLTCGV-----W--- 331
```

Db 1057 AGYARKVQLTVYVPRVFGDLRGLSDOKPVEISVLAGEEVTLPCEVKSLEPPIITWAKE 1116  
QY 332 ----GPTSPK--LMSLLENKEAKVSKREKPVWLNPEAGMOCQLSD--SGOV--LLES 382  
Db 1117 TOLISPSFSPHRTFLPGSMKITETRTS-----DSGMVLCVATNINAGVTOAVKL 1165  
QY 383 NIKVLPWTST-PVHPRASA-----LPAPPTGSALP 411  
Db 1166 NVHVPKIQRGPKHLKQVGVQGVDFPCNAQGTPLP 1200

RESULT 71  
US-10-023-634-93  
; Sequence 93, Application US/10023634  
; Publication No. US20030236389A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkete, Richard A  
; APPLICANT: Colman, Steven D  
; APPLICANT: Spylek, Kimberly A  
; APPLICANT: Ballinger, Robert A  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Shenoy, Sureen G  
; APPLICANT: Li, Li  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Zernusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Gueev, Vladimir Y  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Malyanekar, Uriel M  
; APPLICANT: Gunther, Erik  
; APPLICANT: Smithson, Glenda  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Gerlach, Valerie  
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of  
; FILE REFERENCE: 21402-221  
; CURRENT APPLICATION NUMBER: US/10/023,634  
; PRIOR FILING DATE: 2002-06-28  
; PRIOR APPLICATION NUMBER: 60/256,025  
; PRIOR FILING DATE: 2000-12-15  
; PRIOR APPLICATION NUMBER: 60/265,163  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: 60/272,929  
; PRIOR FILING DATE: 2001-03-02  
; PRIOR APPLICATION NUMBER: 60/274,864  
; PRIOR FILING DATE: 2001-03-09  
; PRIOR APPLICATION NUMBER: 60/276,688  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/277,880  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: 60/286,409  
; PRIOR FILING DATE: 2001-04-25  
; PRIOR APPLICATION NUMBER: 60/309,246  
; PRIOR FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: 60/315,600  
; PRIOR FILING DATE: 2001-08-29  
; NUMBER OF SEQ ID NOS: 132  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 93  
; LENGTH: 5636  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (3003)  
; OTHER INFORMATION: wherein Xaa is any amino acid as defined in the  
; OTHER INFORMATION: specification.

FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (3041)  
; OTHER INFORMATION: wherein Xaa is any amino acid as defined in the  
; OTHER INFORMATION: specification.  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (3367)  
; OTHER INFORMATION: wherein Xaa is any amino acid as defined in the  
; OTHER INFORMATION: specification.  
US-10-023-634-93

Query Match 6.2%; Score 143; DB 15; Length 5636;  
Best Local Similarity 20.4%; Pred. No. 0.53;  
Matches 93; Conservative 56; Mismatches 168; Indels 138; Gaps 21;

QY 34 GDIYELTCTASQKKSIOFHKNSNQIKLQNSQFLTKGSKANDRADSRSLMDQGNFP 93  
Db 807 GSNVTLPCYVQGYEPTIKRRRLDNMFISR--PFSVSIQRLRTGA----- 851  
QY LIINKLKIEDSDTYICEVEDQ-----KEEVQLVFGTLA-----NSDTHLLOGSLTLTL 143  
Db 852 LFLINLMASDKGYICEAENQFGKIQSETTVYTLGVAPLIGISPSVANYIBGOQLTIPC 911  
QY 144 ESPPGS-----SPSVQCRSPRGKNIQGGKTLASVSQLEIQDSGTWTC-- 184  
Db 912 TLLAGNPIPRRWIKNSAMLLQNPYITVR-----DSLHIERVOLQDGGEYTCVAS 963  
QY 185 TVLQNKQVFEKIDIVLAFAQKASIVYKKEGROVERSPFLATVEKLTGSGELMQAER 244  
Db 964 NVAGTNNKTTSVVHVLPTIQHQQLSTIEG--IPVTLP-----CKA 1004  
QY 245 ASSSKSWITPDLKNKEVSVRVTQDPPLQWKKLPHLTLPPQALPYAGSGNTLALAEAK 304  
Db 1005 SGNPKPSVIMSKKGEELIS-----TSSAKFSGAGCSLYVSPEG-----EESGEVTCRATNT 1056  
QY 305 TGKLDQEVNLV-----MRATQLQK-----NLTCV-----W-- 331  
Db 1057 AGYARKVQLTVYVPRVFGDLRGLSDOKPVEISVLAGEEVTLPCEVKSLEPPIITWAKE 1116  
QY 332 ----GPTSPK--LMSLLENKEAKVSKREKPVWLNPEAGMOCQLSD--SGOV--LLES 382  
Db 1117 TOLISPSFSPHRTFLPGSMKITETRTS-----DSGMVLCVATNINAGVTOAVKL 1165  
QY 383 NIKVLPWTST-PVHPRASA-----LPAPPTGSALP 411  
Db 1166 NVHVPKIQRGPKHLKQVGVQGVDFPCNAQGTPLP 1200

RESULT 72  
US-10-408-765A-1895  
; Sequence 1895, Application US/10408765A  
; Publication No. US20040101874A1  
; GENERAL INFORMATION:  
; APPLICANT: Ghosh, Soumitra S.  
; APPLICANT: Fahy, Eoin D.  
; APPLICANT: Zhang, Bing  
; APPLICANT: Gibson, Bradford W.  
; APPLICANT: Taylor, Steven W.  
; APPLICANT: Glenn, Gary M.  
; APPLICANT: Warnock, Dale E.  
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
; FILE REFERENCE: 660088,465  
; CURRENT APPLICATION NUMBER: US/10/408,765A  
; PRIOR FILING DATE: 2003-04-04  
; NUMBER OF SEQ ID NOS: 3077  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1895  
; LENGTH: 5636  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:



NAME/KEY: VARIANT  
LOCATION: 3003, 3041, 3367  
OTHER INFORMATION: Xaa = Any Amino Acid  
US-10-408-765A-1895

Query Match 6.2%; Score 143; DB 16; Length 5636;  
Best Local Similarity 20.4%; Pred. No. 0.53;

Matches 93; Conservative 56; Mismatches 168; Indels 138; Gaps 21;

QY 34 GPTVELTCTASOKKSIQFHKNKSNQIKIIGNOGSFLTKGPSKLNDRADSRSLMDQGNFP 93  
DB 807 GSNVTLPCVQGYPEPTIKMRRLDMNPIFSR--PFSVSSISQLRTGA----- 851  
QY 94 LIIKLKIEDSDTYICEVEDO-----KEEVQLLVFGTLA-----NSDTHLQOGSTLT 143  
DB 852 LFIILWASDCKGTICBAENQFKIQSEETVTVTGLVAPLISPSVANVLEGGQTLPC 911  
QY 144 ESPPS-----SPSVQCRSPRKNIGGKTLVSQLELQDSGTWC--- 184  
DB 912 TLNAGNPIPERMINKSNMNLQNPITYRS-----DGLHIERVQDQDGETCVAS 963  
QY 185 TVLQNKVVEFKIDIVLAFQKASSIVYKKEGQVEFSFPLAFVTEKLTGSGELMWQAR 244  
DB 964 NAGTNNKTSYVVVHLPTIQHQOILSTIEG--IPVTLP-----CKA 1004  
QY 245 ASSSKSWITFDLKNKEVSVKVTQDPKQLQMGKKLPLHLTLPOALPOVAGSGLTLALEAK 304  
DB 1005 SGNPSPSVYWSKKGELIS---TSAKFSAGADGSLVYVSPG---EBSGKYCTATNT 1056  
QY 305 TGLKQEVNLVY-----NRATQLOK-----NLTCGV-----W--- 331  
DB 1057 AGYARKVQGLTVYVPRVFGDLRGISQDKPVEISVLAGEVTLPEVSLPRPIITWAKE 1116  
QY 332 ---GPTSPK--LMLSLKLENKAVSKREKPVWLNPEAGMOCILSD-SCGV--LIES 382  
DB 1117 TQLISPFSPRHTFLPSGSKMITETRTS-----DSGMVLCVATNIAQNTVQAVL 1165  
QY 383 NIKVLPTWST-PVHPRASA-----LPAPTSALP 411  
DB 1166 NVHVPKIQGRPKHLKVQGVQGVDPICNAQTFPL 1200

#### RESULT 73

US-10-108-260A-3037  
Sequence 3037; Application US/10108260A  
Publication No. US20040005560A1  
GENERAL INFORMATION:  
APPLICANT: HELIX RESEARCH INSTITUTE  
TITLE OF INVENTION: No. US20040005560A1el full length cDNA  
FILE REFERENCE: H1-A0106  
CURRENT APPLICATION NUMBER: US/10/108,260A  
CURRENT FILING DATE: 2002-03-27  
NUMBER OF SEQ ID NOS: 5458  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 3037  
LENGTH: 956  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-108-260A-3037

Query Match 6.1%; Score 141; DB 15; Length 956;  
Best Local Similarity 25.6%; Pred. No. 0.066;

Matches 83; Conservative 59; Mismatches 128; Indels 54; Gaps 15;

QY 96 INKLKIEDSDTYICEVEDQKEEVQLLVFGTLA--NSDTHLQ--GQSILTLTLESPPGSSP 151  
DB 592 VENLNL---LRACEETKKEEIKKPEPTLQNMVLEHATLNEAGNPLVENVDSNVGSSI 647  
QY 132 SVQCRSPRKNIGGKTLVSQLEL-----QDSGTW--TCTVLQNK--VEFKIDI 199  
DB 648 SKELR---LNKRWRLVSKTQLEWNLPLMIKQDOPFTDNGNLTLSKEEKATVEFSTDM 704  
QY 200 VVLAQKASSIV---YKKEGEQVEFSFPLAFVTEKLTGSGELMWQARASSKSWITP 254

DB 705 SVELPENNVNITKQGEKKEKNEEFTQGLKAKXQVEKIQGVEI--WEAE-----AKSYLDQ 759  
QY 255 D-LKNKEVSVKRV---TQDPKQMGKCLPLHLTLPOALPOVAGSGLTLALEAKTKL 308  
DB 760 DDVVTSMEEISKHLIAKSMFDELMARSEDMLQMDIONISSQESFOHVLTTGLQAKIOEA 819  
QY 309 HQEVNLVVRATQKXNLTCVWGPPTSFKMLSLKLENKAVSKREKPVWLVNPEAGMW 368  
DB 820 KEKQIVNVKLIALLKNTL-----DVSPDLIRLME--SCKELSTYMPRAQQLLG 869  
QY 369 QC-----LLSDSGVLLSESNIKVL 387  
DB 870 QRESGELISGKKEALISNTKSL 893

#### RESULT 74

US-09-925-301-1133  
Sequence 1133; Application US/09925301  
Patent No. US20020052308A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA106  
CURRENT APPLICATION NUMBER: US/09/925,301  
PRIOR FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05882  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1694  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 1133  
LENGTH: 737  
TYPE: PRT  
ORGANISM: Homo sapiens

FEATURE:  
NAME/KEY: SITE  
LOCATION: (1)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (7)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (140)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (186)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (194)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (308)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (534)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (535)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-301-1133

Query Match 6.0%; Score 140.5; DB 9; Length 737;  
Best Local Similarity 18.7%; Pred. No. 0.051;

Matches 96; Conservative 73; Mismatches 215; Indels 129; Gaps 20;

QY 20 PAATQGNKRVVLGGKGDYVELTCTASQKKSIOFHKKNKSNQIKIIGNOGSFLTKGSKLND 79  
DB 182 PSIXNNSNPVEKX-DAVAFCEBEVQNTTYLWVWNGSLPV----- 222  
QY 80 ADSRRSLMDQGNFPLIINKLKIEDSDTYICEV-----DQKEEVQLLVF-----GLTAN-S 129

```

Db      223  --SPRLQSLNGNMTLLSVKRNDA5YECEI0NPASANRSDPYTLNLYXGPDGPTTSPS 280
Qy      130  DTHLLQGSGLTLESPGSSPVOCR--SPRGKNIQGGKTLVSQLELQDSGTWTCYVLQ 188
Db      281  KANRPGENLMSCHA--ASNPPA0YSWFKNGTFQOSTOELPIPIVITVNSG5TYCOAHN 338
Qy      189  NQKKV-EFKIDIVLAFQKASSIVYKKEQVEFSFPLATVE--KLTSGELM0AERAS 246
Db      339  SDTGLNRTTWTITVYAPPKPFITSNNSNPVEDDANALCEPEIQNTTYLMMVNNQSL 398
Qy      247  SSKSMTTDLKXKEYSVKRYTODP-----KLQGGKULPLHL-----TLPOA 287
Db      399  PVSPRLQSLNDRITLLSVTRNDVGRPYECGIONELSYDHSDPYTLNLYXGPDGPTTSPS 458
Qy      288  LPOYAGSGLTLALAEKT-----GKLHOEV-NLYVMRATQLOKUL-TCEWGPPT 334
Db      459  YTYRPGVNLISCHASNPAPQYSWLDGNIQDHTQELFISNTEKNSGLYTCOANNSA 518
Qy      335  S-----PKMLSLKLENKEAKVSKREKPVWVLPNPEAGMOCCLSDSGOV 378
Db      519  SGHRTTVKTTIVSAXXKPSIS---SNNSKPVEDKDAVAFCEPEAQNTTYLMMVNGQ- 574
Qy      379  LLESNIKVLPTWSTFVHPRASALPAPPT-----GSLPDPQTA 416
Db      575  -----SLPVSRLQSLNGNRTLTLENTVTRNDARAYVCGIQNSVANSRSDPYTL 622
Qy      417  SAL--PDPPASALPAPALAVISFLGLGLGVAC 447
Db      623  DVLGPDPTIISPPDS-----SYLSGANLNLISC 650

```

## RESULT 75

```

US-10-311-823-16
; Sequence 16, Application US/10311823
; Publication NO. US20040116683A1
; GENERAL INFORMATION:
; APPLICANT: Bryan J. Boyle
; APPLICANT: Nancy Mize
; APPLICANT: Matthew Arterburn
; APPLICANT: Y. Tom Tang
; APPLICANT: George Yeung
; APPLICANT: Ping Zhou
; APPLICANT: Chenghua Liu
; APPLICANT: Vinod Asundi
; APPLICANT: Radoje T. Drmanac
; APPLICANT: Meng-Yun Wang
; APPLICANT: Lichuan Chen
; APPLICANT: Yea-Huey Yang
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NEUROTROPIN-LIKE
; FILE REFERENCE: HYS-17CIP/US
; CURRENT APPLICATION NUMBER: US/10/311,823
; PRIOR FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: PCT/US01/03651
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/632,085
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-311-823-16

```

```

Query Match      6.0%; Score 139.5; DB 16; Length 374;
Best Local Similarity 23.5%; Pred. No. 0.024;
Matches 82; Conservative 54; Mismatches 136; Indels 77; Gaps 18;

```

```

Qy      36  TVELTCTASQKSI0FHWKNSNOIKILNGSFLTKGPSKLANDRADSRRLMDQGNPLI 95
Db      55  TVFLRCLTVNSNPAPRIFWKRGS-----TISHQDNG-VDIYEELTYQGETKVL 102
Qy      96  -INKLKIEDSDTYICEVEDQK--EVQLLVFGLTANSDTHLLQGSGLTLESPPGSSP 151
Db      103  KKLNLAPQDYASATCGVSVANVCGIPDKAITFFRLTYTTAPPAK-LSVNTLLVNPGENV 161
Qy      152  SYVC-----RSPRGKNIQGGKTLVSQLELQDSGTWTCYVLQ----- 169
Db      162  TVQCLLTGDPPLPOLQMSHGPPPLGALAQGG-TLSISVQARSDSYNCTATNNVGNP 220
Qy      190  QKVEFKIDIVLAFQKASSIVYK--KEGEQVEFSFPL-----AFTVEKLTGSGELW 239
Db      221  AKKT---VNLVMSKNATQITQIPDIYKESENIQLQGDLLSCHVAVPQEKTYQ---W 274
Qy      240  WQ-ARASSSKSKSMTTDPDLKXKEYSVKRYTODPDLQGGKULPLHL-----TLPOA- 287
Db      275  FKNGKPRM5KRL-----VTENDELPAVTSLSLEL-----IDLHFSQGYTLCVASFPGAP 326
Qy      288  LPOYAGSGLTLALAEKTGKLEHENVLYVMRATQLOKULCEVWGPTSP 336
Db      327  VPDLSVEVNISSETVPTTISVPKRAVTVREGS-PAELQCEVRGKPRP 374

```

## RESULT 76

```

US-10-311-823-12
; Sequence 12, Application US/10311823
; Publication NO. US20040116683A1
; GENERAL INFORMATION:
; APPLICANT: Bryan J. Boyle
; APPLICANT: Nancy Mize
; APPLICANT: Matthew Arterburn
; APPLICANT: Y. Tom Tang
; APPLICANT: George Yeung
; APPLICANT: Ping Zhou
; APPLICANT: Chenghua Liu
; APPLICANT: Vinod Asundi
; APPLICANT: Radoje T. Drmanac
; APPLICANT: Meng-Yun Wang
; APPLICANT: Lichuan Chen
; APPLICANT: Yea-Huey Yang
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NEUROTROPIN-LIKE
; FILE REFERENCE: HYS-17CIP/US
; CURRENT APPLICATION NUMBER: US/10/311,823
; PRIOR FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: PCT/US01/03651
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/632,085
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-311-823-12

```

```

Query Match      6.0%; Score 138.5; DB 16; Length 442;
Best Local Similarity 25.6%; Pred. No. 0.036;
Matches 60; Conservative 36; Mismatches 85; Indels 53; Gaps 11;
Qy      36  TVELTCTASQKSI0FHWKNSNOIKILNGSFLTKGPSKLANDRADSRRLMDQGNPLI 95
Db      136  TVFLRCLTVNSNPAPRIFWKRGS-----TISHQDNG-VDIYEELTYQGETKVL 183
Qy      96  -INKLKIEDSDTYICEVEDQK--EVQLLVFGLTANSDTHLLQGSGLTLESPPGSSP 151

```

```
Db      184 KLNKLRPODYASVTCQVSVRVNCGIPDKAITFRLTNTTAPALK-LSVNETLVNPGENV 242
Qy      152 SVQC-----RSPRGKNIQGGKTLVSQLELOSGTWTCTVLON----- 189
Db      243 TVQCCLLTGDDPLPQLQWHSHPPLPLGALAQGG-TLSIPSVQARDSGYNCATATNVGNP 301
Qy      190 OKKVEFKIDIVLAFQKASSIVYK--KEGEQVEFSFPL-----AFTVEKLT 233
Db      302 AKKT---VNLVLRSMKMAITFQITPVIKESENIQLOGDLKLSCHDAVQGEKVT 352

RESULT 77
US-10-311-823-4
; Sequence 4, Application US/10311823
; Publication No. US20040116683A1
; GENERAL INFORMATION:
; APPLICANT: Bryan J. Boyle
; APPLICANT: Nancy Mize
; APPLICANT: Matthew Arterburn
; APPLICANT: Y. Tom Tang
; APPLICANT: George Yeung
; APPLICANT: Ping Zhou
; APPLICANT: Chenghua Liu
; APPLICANT: Vinod Asundi
; APPLICANT: Radoje T. Drmanac
; APPLICANT: Meng-Yun Wang
; APPLICANT: Lichuan Chen
; APPLICANT: Yea-Huey Yang
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NEUROTROPIN-LIKE
; FILE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES
; FILE REFERENCE: HYS-17CIP/US
; CURRENT APPLICATION NUMBER: US/10/311,823
; PRIOR FILING DATE: 2003-09-29
; PRIOR APPLICATION NUMBER: PCT/US01/03651
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 09/632,085
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-311-823-4

Query Match      6.0%; Score 138.5; DB 16; Length 458;
Beet Local Similarity 25.6%; Pred. No. 0.038;
Matches 60; Conservative 36; Mismatches 85; Indels 53; Gaps 11;

Qy      36 TVELCTASQKKSIOFHMKNNOIKILGNQGSFLTGKPSKLNDRADRSRLSDQGNFPLI 95
Db      152 TVFLACTVNSNPAPFIWGRGSD-----TLHSQDNG-VDIYPLTYQGETKYL 199
Qy      96 -IKNLIKIEDSDTYICEVEDQKE--EVQLLVFGLTANSDTHLLQGSITLTLESPPGSSP 151
Db      200 KLNKLRPODYASVTCQVSVRVNCGIPDKAITFRLTNTTAPALK-LSVNETLVNPGENV 258
Qy      152 SVQC-----RSPRGKNIQGGKTLVSQLELOSGTWTCTVLON----- 189
Db      259 TVQCCLLTGDDPLPQLQWHSHPPLPLGALAQGG-TLSIPSVQARDSGYNCATATNVGNP 317
Qy      190 OKKVEFKIDIVLAFQKASSIVYK--KEGEQVEFSFPL-----AFTVEKLT 233
Db      318 AKKT---VNLVLRSMKMAITFQITPVIKESENIQLOGDLKLSCHDAVQGEKVT 368

RESULT 78
US-09-905-129-21
; Sequence 21, Application US/09905129
```

```
; Patent No. US20020137705A1
; GENERAL INFORMATION:
; APPLICANT: Bimat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THERE
; FILE REFERENCE: 540579-2007.2
; CURRENT APPLICATION NUMBER: US/09/905,129
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/207,821
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: 60/084,944
; PRIOR FILING DATE: 1998-05-11
; PRIOR APPLICATION NUMBER: 60/085,673
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 2828
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-905-129-21

Query Match      5.9%; Score 138; DB 9; Length 2828;
Beet Local Similarity 19.4%; Pred. No. 0.5;
Matches 103; Conservative 63; Mismatches 180; Indels 184; Gaps 21;

Qy      10 LLLVQLALLPAPATQGNKV--VLGKGGDTVELTCTASQKKSIOFHMKNNOIKILGNQGS 67
Db      2036 LAIRLHVALLPVPVHQEKLENIISLPGLSIHICHTAKAAPLPSVRW-----VLG----- 2084
Qy      68 FLTKPSKLNDRADRSRLMDQGN--FP---LIINKLIEDSDTYICE--VEDQKEE 117
Db      2085 -----DGTQIRPQFLHGNLFVFPNGTLYIRNLAPDSGRYECAANLVGSART 2134
Qy      118 VOLLVFGLTANS-----DTHLLOGSLTLTLESPPGSSPSVQCRSPRGKNIQ----- 164
Db      2135 VOLNVORAAANARITGTSPPRTDVRYGGLTKLDCSASGDDPMPRIIMRLPSKRMIDALFSF 2194
Qy      165 -----GKTLVSQLELOSGTWTCTVLONQ--KKVEFKIDIVLAFQKASSIVYK 214
Db      2195 DSRIVFANGTLVNVKSVTDKAGDYLC-VARNKVGDDVVLKVDVW---KPAKIEHKE 2249
Qy      215 EGE-----QVEFSFPLAFTVEKLTGSGELMWAERASSSKSWIT 253
Db      2250 ENDHKVPYGGDLKVDCVATGHPNBEISWISLPSGSLVNSFMQSD-----DSGGRTKRYV 2303
Qy      254 FD---LKNKEVSVKRVYQDPLQMGKKLPLHLTLPLQALPQVAGSGNLTLLALEAKTGKXHQ 310
Db      2304 FNNGLTYFNEVGMR-----BDDYTCFANOVGKDEM 2335
Qy      311 EVNLVVMKATOLQKULT-----CEWAGPSPKMLSLKLENKEAKVSKRE 355
Db      2336 RVRKVVVTAAPATINKTTLAVQVVPYGDVYVTAACEKAGKPMKVTW-LSPTNKVLPITSEK 2394
Qy      356 KPVM-----VLNPEAGMQCLSDSG-----QVLESNIVKLPWTSPVHPRASAL 401
Db      2395 YQIVQDGTLLIQAKQSDSGNYTCIVNSAGEDRKTVIHNVO----- 2438
Qy      402 PAPPTGSLPDPQT-----ASLPPPPASALPALAV 434
Db      2439 --PPKINGNPNPITTVREIAAGSRKLIIDCKAEGIPTPRVILMAPPEGVVL 2486

RESULT 79
US-09-991-630-21
; Sequence 21, Application US/09991630
; Patent No. US200201514A1
; GENERAL INFORMATION:
; APPLICANT: Bimat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THERE
```

```

; FILE REFERENCE: 540579-2007.3
; CURRENT APPLICATION NUMBER: US/09/991,630
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: 09/905,129
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 09/729,485
; PRIOR FILING DATE: 2000-12-04
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 2828
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-991-630-21

Query Match      5.9%; Score 138; DB 9; Length 2828;
Best Local Similarity 19.4%; Pred. No. 0.5;
Matches 103; Conservative 63; Mismatches 180; Indels 184; Gaps 21;

QY 10 LLLVLOALLPAAVQGNKV--VLGKKGDTVELCTASQKSIQFHWKNSNQIKILGNQGS 67
DB 2036 LAIRLHVALLPPIVHQQEKLNIISLPPGLSIHICTAKAAPLPSVRW-----VLG---- 2084
QY 68 FLTGPSKLNDRADSRSLMDQGN---FP---LIINKLKIEDSDTYICE---VEDQKEE 117
DB 2085 -----DGTQIRPSQFLHGNLFVFPNGTLTYIRNLAPDSGRYECVAAVLGSAART 2134
QY 118 VOLLVFGLTANS-----DTHLLOQSLLTLTLESPGSSPSVQCRSPKGNIQ----- 164
DB 2135 VOLNVQRAAANAARITGTSPPRTDVRVYGGTLKIDCSASGDPWPRILMRLPSKRMIDALFSF 2194
QY 165 -----CGKTLVSQLELQDSGTWTCTVLONQ---KVEFKIDIYVLARQKASSIYKK 214
DB 2195 DSRIRVPANGTLVAVKSDTKDAGDYLC-VARNKVGDDYVVLKVDVWV---KPAKIEHKE 2249
QY 215 EGE-----QVEFSFPLAFTVEKLTGSGELMWQAEARASSSKSWIT 253
DB 2250 ENDHKVFYGGDLKVDCAVATGLPNPEISWSLSDGSLVNSFMQSD-----DSGRTKRYV 2303
QY 254 FD---LKNKEVSVKRVTPDKLQMGKKLPLHLTLPOALPOYAGSGNLTALAEATGKLIHQ 310
DB 2304 FNNGTLYFNEVGMBE-----EGDYTCFAENQVQKDEM 2335
QY 311 EVNLVVMATQLOKRLT-----CEWGGTSPKMLMSLKEHKEKAVSGRE 355
DB 2336 RVRKVVVTAAPATIRNKTYLAVQVPYGDVTVVACEAKGSPMKVTV-LSPTRKVIPTSEK 2394
QY 356 KPVW-----VLNPEAGMOCILSDSG-----QVLLSENIKVLPTWSTPVHPRASAL 401
DB 2395 VOIYQDGTGLIQKQKRSDSGNYTCLVNSAGEDRTKTVIHNVQ----- 2438
QY 402 PAPPTGSLPDPQT-----ASALPDPPEASALPALAV 434
DB 2439 --PKINGNPNPITTVREIAAGSRKLIIDCKAEGIPTRVILMAPEEGVVL 2486

RESULT 80
US-10-072-012-607
; Sequence 607, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchenev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zetnusen, Bryan
; APPLICANT: Patuturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Beha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Raestelli, Luca

```

```

; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosee, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 607
; LENGTH: 2828
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-607

Query Match      5.9%; Score 138; DB 12; Length 2828;
Best Local Similarity 19.4%; Pred. No. 0.5;
Matches 103; Conservative 63; Mismatches 180; Indels 184; Gaps 21;

QY 10 LLLVLOALLPAAVQGNKV--VLGKKGDTVELCTASQKSIQFHWKNSNQIKILGNQGS 67
DB 2036 LAIRLHVALLPPIVHQQEKLNIISLPPGLSIHICTAKAAPLPSVRW-----VLG---- 2084
QY 68 FLTGPSKLNDRADSRSLMDQGN---FP---LIINKLKIEDSDTYICE---VEDQKEE 117
DB 2085 -----DGTQIRPSQFLHGNLFVFPNGTLTYIRNLAPDSGRYECVAAVLGSAART 2134
QY 118 VOLLVFGLTANS-----DTHLLOQSLLTLTLESPGSSPSVQCRSPKGNIQ----- 164
DB 2135 VOLNVQRAAANAARITGTSPPRTDVRVYGGTLKIDCSASGDPWPRILMRLPSKRMIDALFSF 2194
QY 165 -----CGKTLVSQLELQDSGTWTCTVLONQ---KVEFKIDIYVLARQKASSIYKK 214
DB 2195 DSRIRVPANGTLVAVKSDTKDAGDYLC-VARNKVGDDYVVLKVDVWV---KPAKIEHKE 2249
QY 215 EGE-----QVEFSFPLAFTVEKLTGSGELMWQAEARASSSKSWIT 253
DB 2250 ENDHKVFYGGDLKVDCAVATGLPNPEISWSLSDGSLVNSFMQSD-----DSGRTKRYV 2303
QY 254 FD---LKNKEVSVKRVTPDKLQMGKKLPLHLTLPOALPOYAGSGNLTALAEATGKLIHQ 310
DB 2304 FNNGTLYFNEVGMBE-----EGDYTCFAENQVQKDEM 2335

```



```

Best Local Similarity 19.4%; Pred. No. 0.5;
Matches 103; Conservative 63; Mismatches 180; Indels 184; Gaps 21;

Qy 10 LLLVQLALPPAATGKNV--VLGKGGDTVELTCTASQKSIQFHMKNNSQIKILNGGS 67
Db LAIRLHVAAALPPVHQEKLENIISLPPGLSIHICTAKAAPLPSVRW-----VLG----- 2084

Qy 68 FLTKGPKLNDRAISRSLMDGN--FP--LIINKLKIEDSDTYICE---VEDQKEE 117
Db -DGTQIRPSQFLHGNLFVFPNGTLYIRNLAPDGSRGVECVANILVGSARRT 2134

Qy 118 VOLLVFGILTANS-----DTHLLOGSITLTLESPPGSSPSVQCRSPRGKNIQ----- 164
Db VOLNVQRAAANAARITGTSPPRTDVRYGGTILKDCSASGDWPRIILWRLPSKRMIDALFSF 2194

Qy 165 -----GKTLVSQLELDSDGTWCTVLQNO---KVEFKIDIYVLAFOKASSIYKK 214
Db DSRIKVPANGTLVVKSVTDKADGYLC-VARNKVGDDYVVLKVDVVM---KPAKIEHKE 2249

Qy 215 EGE-----QVEFSPLAFTVEKLTGSGELMWQARRASSSKSWIT 253
Db ENDHKVFYGGDLKVDCAVATGLPNEPISWSLPGSLVNSFMQSD-----DSGRTKRYV 2303

Qy 254 FD---LKNKEYSVKRVTDPKLQMGKKLPLHLTLPOALPYAGSGNLTLEAKTGKLNQ 310
Db FNNGTLYFNEVGMBE-----EGDYTCFAENQVGDDEM 2335

Qy 311 EVNLVVMRATQLOKNTL-----CEWGPPTS PKMLSLIKENKEAKVSKRE 355
Db RVRKVVTAATATIRNKTYLAVQVPYGDVVTVACEKAGEPMKVTW--LSPTNKV IPTSEK 2394

Qy 356 KPVM-----VLNPEAGMOCILSDSG---QVLESNIKVLPTWSTPVHPRASAL 401
Db YQIYODGTLTIQKQRSDSGNYTCLVNSAGEDRKTWIIHVNQ----- 2438

Qy 402 PAPPTGSALPDPOT-----ASALPDPASALPALAV 434
Db 2439 --PKINGNPNPITTYREIAAGSRKLDCKAEGIPPRVLMAPFEGVVL 2486

RESULT 83
US-09-802-318-21
; Sequence 21, Application US/09802318
; Publication No. US20020086825a1
; GENERAL INFORMATION:
; APPLICANT: Elnat, et al
; TITLE OF INVENTION: GENES ASSOCIATED WITH MECHANICAL STRESS, EXPRESSION PRODUCTS THERE
; FILE REFERENCE: 540579-2007.1
; CURRENT APPLICATION NUMBER: US/09/802,318
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 09/632,862
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/207,821
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 2828
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-802-318-21

Query Match 5.9%; Score 138; DB 12; Length 2828;
Best Local Similarity 19.4%; Pred. No. 0.5;
Matches 103; Conservative 63; Mismatches 180; Indels 184; Gaps 21;

Qy 10 LLLVQLALPPAATGKNV--VLGKGGDTVELTCTASQKSIQFHMKNNSQIKILNGGS 67
Db LAIRLHVAAALPPVHQEKLENIISLPPGLSIHICTAKAAPLPSVRW-----VLG----- 2084

Qy 68 FLTKGPKLNDRAISRSLMDGN--FP--LIINKLKIEDSDTYICE---VEDQKEE 117

```

```

Db 2085 -----DGTQIRPSQFLHGNLFVFPNGTLYIRNLAPDGSRGVECVANILVGSARRT 2134
Qy 118 VOLLVFGILTANS-----DTHLLOGSITLTLESPPGSSPSVQCRSPRGKNIQ----- 164
Db VOLNVQRAAANAARITGTSPPRTDVRYGGTILKDCSASGDWPRIILWRLPSKRMIDALFSF 2194

Qy 165 -----GKTLVSQLELDSDGTWCTVLQNO---KVEFKIDIYVLAFOKASSIYKK 214
Db DSRIKVPANGTLVVKSVTDKADGYLC-VARNKVGDDYVVLKVDVVM---KPAKIEHKE 2249

Qy 215 EGE-----QVEFSPLAFTVEKLTGSGELMWQARRASSSKSWIT 253
Db ENDHKVFYGGDLKVDCAVATGLPNEPISWSLPGSLVNSFMQSD-----DSGRTKRYV 2303

Qy 254 FD---LKNKEYSVKRVTDPKLQMGKKLPLHLTLPOALPYAGSGNLTLEAKTGKLNQ 310
Db FNNGTLYFNEVGMBE-----EGDYTCFAENQVGDDEM 2335

Qy 311 EVNLVVMRATQLOKNTL-----CEWGPPTS PKMLSLIKENKEAKVSKRE 355
Db RVRKVVTAATATIRNKTYLAVQVPYGDVVTVACEKAGEPMKVTW--LSPTNKV IPTSEK 2394

Qy 356 KPVM-----VLNPEAGMOCILSDSG---QVLESNIKVLPTWSTPVHPRASAL 401
Db 2395 YQIYODGTLTIQKQRSDSGNYTCLVNSAGEDRKTWIIHVNQ----- 2438

Qy 402 PAPPTGSALPDPOT-----ASALPDPASALPALAV 434
Db 2439 --PKINGNPNPITTYREIAAGSRKLDCKAEGIPPRVLMAPFEGVVL 2486

RESULT 84
US-10-176-847-54
; Sequence 54, Application US/10176847
; Publication No. US20030068636a1
; GENERAL INFORMATION:
; APPLICANT: Velby, Pectier Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/10/176,847
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54
; LENGTH: 2828
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-176-847-54

Query Match 5.9%; Score 138; DB 14; Length 2828;
Best Local Similarity 19.4%; Pred. No. 0.5;
Matches 103; Conservative 63; Mismatches 180; Indels 184; Gaps 21;

Qy 10 LLLVQLALPPAATGKNV--VLGKGGDTVELTCTASQKSIQFHMKNNSQIKILNGGS 67
Db LAIRLHVAAALPPVHQEKLENIISLPPGLSIHICTAKAAPLPSVRW-----VLG----- 2084

Qy 68 FLTKGPKLNDRAISRSLMDGN--FP--LIINKLKIEDSDTYICE---VEDQKEE 117
Db -DGTQIRPSQFLHGNLFVFPNGTLYIRNLAPDGSRGVECVANILVGSARRT 2134

Qy 118 VOLLVFGILTANS-----DTHLLOGSITLTLESPPGSSPSVQCRSPRGKNIQ----- 164
Db VOLNVQRAAANAARITGTSPPRTDVRYGGTILKDCSASGDWPRIILWRLPSKRMIDALFSF 2194

Qy 165 -----GKTLVSQLELDSDGTWCTVLQNO---KVEFKIDIYVLAFOKASSIYKK 214
Db DSRIKVPANGTLVVKSVTDKADGYLC-VARNKVGDDYVVLKVDVVM---KPAKIEHKE 2249

Qy 215 EGE-----QVEFSPLAFTVEKLTGSGELMWQARRASSSKSWIT 253
Db ENDHKVFYGGDLKVDCAVATGLPNEPISWSLPGSLVNSFMQSD-----DSGRTKRYV 2303

Qy 254 FD---LKNKEYSVKRVTDPKLQMGKKLPLHLTLPOALPYAGSGNLTLEAKTGKLNQ 310
Db FNNGTLYFNEVGMBE-----EGDYTCFAENQVGDDEM 2335

Qy 311 EVNLVVMRATQLOKNTL-----CEWGPPTS PKMLSLIKENKEAKVSKRE 355
Db RVRKVVTAATATIRNKTYLAVQVPYGDVVTVACEKAGEPMKVTW--LSPTNKV IPTSEK 2394

Qy 356 KPVM-----VLNPEAGMOCILSDSG---QVLESNIKVLPTWSTPVHPRASAL 401
Db 2395 YQIYODGTLTIQKQRSDSGNYTCLVNSAGEDRKTWIIHVNQ----- 2438

Qy 402 PAPPTGSALPDPOT-----ASALPDPASALPALAV 434
Db 2439 --PKINGNPNPITTYREIAAGSRKLDCKAEGIPPRVLMAPFEGVVL 2486

```

Db 2250 ENDHKVFGYGDLLKVDCAVATGPNPEISWGLPDGSLVNSFMQSD-----DSGGRTRKYV 2303  
QY 254 FD---LKKKEVSVKRVTDQPKLQMGKKLPLHLTLPOALPOYAGSGNLTALFAKTKGLHQ 310  
Db 2304 FNNGLTYNEVGMRE-----EGDYTCFPAENOVGKDEM 2335  
QY 311 EVNLVVMRATQLOKNTL-----CEWVGPTSPKMLSLKLENKEAKVSRE 355  
Db 2336 RRVKVVVYAPATIRKTKTYLAVQVPYGVVTVACAKGEMPKVTW-LSPTNKVITPTSSEK 2394  
QY 356 KPWV-----VLNPEAGMWQCLLSDSG-----QVLESNIKVLPTWSTFVHPRASAL 401  
Db 2395 YQIYODGTLTIQAKQRSDSGNYTCLVRNSAGEDRKTWVIHNVQ----- 2438  
QY 402 PAPPTGSALPDPQT-----ASALPDPASALPALAV 434  
Db 2439 --PKINGNPNPITTVREIAAGSRKLLDCKAEGIPTPRVLMAPEGVVL 2486

RESULT 85  
US-10-177-293-110  
; Sequence 110, Application US/10177293  
; Publication No. US20030124128A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Glatt, Karen  
; APPLICANT: Zhao, Xumei  
; APPLICANT: Gannavarpu, Manjula  
; APPLICANT: Kametkar, Shubhangi  
; APPLICANT: Mertens, Maureen  
; APPLICANT: Myer, Vic  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Hoersch, Sebaectian  
; APPLICANT: Monahan, John  
; APPLICANT: Meyers, Rachel E.  
; APPLICANT: Baer Jr., Robert C.  
; APPLICANT: Horrobagyl, Gabriel N.  
; APPLICANT: Pustel, Lajos  
; APPLICANT: Meric, Funda  
; APPLICANT: Sahin, Aysegul B.  
; APPLICANT: Mills, Gordon B.  
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,  
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER  
; FILE REFERENCE: MRI-038  
; CURRENT APPLICATION NUMBER: US/10/177,293  
; PRIOR FILING DATE: 2002-06-21  
; PRIOR APPLICATION NUMBER: US 60/299,887  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 60/301,572  
; PRIOR FILING DATE: 2001-06-27  
; PRIOR APPLICATION NUMBER: US 60/306,501  
; PRIOR FILING DATE: 2001-07-18  
; PRIOR APPLICATION NUMBER: US 60/325,002  
; PRIOR FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US 60/362,585  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx  
; PRIOR FILING DATE: 2002-05-14  
; NUMBER OF SEQ ID NOS: 506  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 110  
; LENGTH: 2828  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-177-293-110

Query Match 5.9%; Score 138; DB 14; Length 2828;  
Best Local Similarity 19.4%; Pred. No. 0.5; Indels 184; Gaps 21;  
Matches 103; Conservative 63; Mismatches 180; Indels 184; Gaps 21;  
QY 10 LLLVLQALLPAAQGNKV--VLGKKGDTVELTCTASQKKSIGFHWKNSNQIKILGNQGS 67

Db 2036 LAIRLHVAALPPVHIOEKLNIISLPPGSLIHICTAKAAPLPSVRW-----VLG----- 2084  
QY 68 FLTKGPSCLNDRABSRRLMPOGN---FP---LIINKLKIDSDPTICE---VEDQKEE 117  
Db 2085 -----DGTQIRPQFLHGNLFPFPNGTLYIRNLAPPDSGRYECVAANLVGSART 2134  
QY 118 VOLVFGTLTANS-----DTHLQGSLLTLTLESPPGSSPSVQCRSPRGKNIQ----- 164  
Db 2135 VOLNVQRAAANARITGTSPPRTDVRVGGTLLKDCSASGDPMPRIIMRLPSKRMIDALNSF 2194  
QY 165 -----GKTLVSQLELQDSGTWCTVLQNO---KVEFKIDIYVLAFOKASIVYKK 214  
Db 2195 DSRIRKFPANGTLVVVGSVTDKQAGDYLC-VARNKQDDYVVLKVDVW---KPAKIEHKE 2249  
QY 215 EGE-----QVERSPFLATVEYELTSSGELIMQAEASSSKSIIT 253  
Db 2250 ENDHKVFGYGDLLKVDCAVATGPNPEISWGLPDGSLVNSFMQSD-----DSGGRTRKYV 2303  
QY 254 FD---LKKKEVSVKRVTDQPKLQMGKKLPLHLTLPOALPOYAGSGNLTALFAKTKGLHQ 310  
Db 2304 FNNGLTYNEVGMRE-----EGDYTCFPAENOVGKDEM 2335  
QY 311 EVNLVVMRATQLOKNTL-----CEWVGPTSPKMLSLKLENKEAKVSRE 355  
Db 2336 RRVKVVVYAPATIRKTKTYLAVQVPYGVVTVACAKGEMPKVTW-LSPTNKVITPTSSEK 2394  
QY 356 KPWV-----VLNPEAGMWQCLLSDSG-----QVLESNIKVLPTWSTFVHPRASAL 401  
Db 2395 YQIYODGTLTIQAKQRSDSGNYTCLVRNSAGEDRKTWVIHNVQ----- 2438  
QY 402 PAPPTGSALPDPQT-----ASALPDPASALPALAV 434  
Db 2439 --PKINGNPNPITTVREIAAGSRKLLDCKAEGIPTPRVLMAPEGVVL 2486

RESULT 86  
US-10-301-822-49  
; Sequence 49, Application US/10301822  
; Publication No. US20030148410A1  
; GENERAL INFORMATION:  
; APPLICANT: Millennium Pharmaceuticals, Inc.  
; APPLICANT: Berger, Allison  
; APPLICANT: Guillemette, Tracy L.  
; APPLICANT: Kametkar, Shubhangi  
; APPLICANT: Schlegel, Robert  
; APPLICANT: Monahan, John E.  
; APPLICANT: Bursart, Lawrence N.  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND  
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; TITLE OF INVENTION: THERAPY OF COLON CANCER  
; FILE REFERENCE: MPM01-029P22RM  
; CURRENT APPLICATION NUMBER: US/10/301,822  
; PRIOR FILING DATE: 2002-11-21  
; PRIOR APPLICATION NUMBER: US 60/339,971  
; PRIOR FILING DATE: 2001-12-10  
; PRIOR APPLICATION NUMBER: US 60/361,978  
; PRIOR FILING DATE: 2002-03-05  
; PRIOR APPLICATION NUMBER: US 60/381,988  
; PRIOR FILING DATE: 2002-05-20  
; NUMBER OF SEQ ID NOS: 228  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 49  
; LENGTH: 2828  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
US-10-301-822-49

Query Match 5.9%; Score 138; DB 14; Length 2828;  
Best Local Similarity 19.4%; Pred. No. 0.5;  
Matches 103; Conservative 63; Mismatches 180; Indels 184; Gaps 21;  
QY 10 LLLVLQALLPAAQGNKV--VLGKKGDTVELTCTASQKKSIGFHWKNSNQIKILGNQGS 67

```

Db      2036 LAIRLHVAAALPPVHIOEKLENIISLPPGLSIHIHCTAKAAPLPSVRW-----VLG----- 2084
Qy      68 FLTGSPSKLNDRAISRSLMDQGN---FP---LIINKLIEDSPYICE---VEDQKEE 117
Db      2085 -----DGTQIRPSQFLHGNLFVFPNGTLTYIRNLAPDSGRYECVAAANLVGSAART 2134
Qy      118 VOLLVFGLTANS-----DTHLLOGQSITLTLESPPGSSPSVQCRSPRKNIO----- 164
Db      2135 VOLNVQRAAANARITGTSPPRTDVRYGGLTKLDCSASGDPRIILMLPSKRMIDALFSF 2194
Qy      165 -----GGKTLVSQLELDQSGTWCTVLONQ---KVEFKIDIYVLAFQKASSIYKK 214
Db      2195 DSRIRKVPANGTLVVKSVTDKADGDLIC-VARNKVGDYVVLKVDVVM---KPAKIEHKE 2249
Qy      215 EGE-----QVERFPLATFYEKLTGSGELMWQABRASSSKSMIT 253
Db      2250 ENDRKVFYGGDLKVDCAVATGLPNPEISWSLPDGLVNSFWQSD-----DSGGRTKRYV 2303
Qy      254 FD---LKNKEVSVKRVTDPKLOMGKKLPLHLTLPOALPOYAGSGNLTLEAKTGKHLQ 310
Db      2304 FNNGTLYFNEVGME-----EGDYTCFAENQVGDDEM 2335
Qy      311 EVNLVVRATOLQKNLT-----CEVWGPTSPKMLSLKLENKEAKSVKRE 355
Db      2336 RVRKVVTAPATIRNKTYLAQVPEYGDVTVVACEAKGEMPKVTW-LSPTNKVIPTSSSEK 2394
Qy      356 KPVW-----VLNPEAGMWOCILSDSG---QVLESNIKVLPTWSPVHPRASAL 401
Db      2395 YQIYQDGTLLIQKQKSDSGNYTCLVRNSAGEDRKTYWHVNVQ----- 2438
Qy      402 PAPPTGSALPDPOT-----ASALPDPPEASALPALAV 434
Db      2439 --PPKINGNPNPITTVREIAAGSRKLLDCKAEGIPFPVLMAPPEGVVL 2486

```

```

RESULT 87
US-10-032-189-126
; Sequence 126, Application US/10032189
; Publication No. US20030170630A1
; GENERAL INFORMATION:
; APPLICANT: Alsedbrook II, John P
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zernusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Groose, William M
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimketers, Richard A
; APPLICANT: Groose, William M
; APPLICANT: Szekeres, Edward S
; APPLICANT: Vermet, Corine A.M.
; APPLICANT: Li, Li
; APPLICANT: Caeman, Stacie J
; APPLICANT: Boldog, Ferenc L
; APPLICANT: Gorman, Linda
; APPLICANT: Gangolli, Esha A
; APPLICANT: Fernandez, Elina R
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Gunther, Erik
; APPLICANT: Miller, Isabelle
; APPLICANT: Sciore, Paul
; APPLICANT: Ellerman, Karen
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glenda
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-228
; CURRENT APPLICATION NUMBER: US/10/032,189
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/257,495

```

```

; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/258,171
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/269,940
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: 60/274,192
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/277,826
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/279,840
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/282,981
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/283,656
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/309,247
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 126
; LENGTH: 2828
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-032-189-126

```

```

Query Match 5.9%; Score 138; DB 14; Length 2828;
Best Local Similarity 19.4%; Pred. No. 0.5;
Matches 103; Conservative 63; Mismatches 180; Indels 184; Gaps 21;

```

```

Qy      10 LLLVQLALIPATQGNKV-VLGKKGDTVELCTASQKKSIGFHWGNSNQIILNGQS 67
Db      2036 LAIRLHVAAALPPVHIOEKLENIISLPPGLSIHIHCTAKAAPLPSVRW-----VLG----- 2084
Qy      68 FLTGSPSKLNDRAISRSLMDQGN---FP---LIINKLIEDSPYICE---VEDQKEE 117
Db      2085 -----DGTQIRPSQFLHGNLFVFPNGTLTYIRNLAPKDSGRYECVAAANLVGSAART 2134
Qy      118 VOLLVFGLTANS-----DTHLLOGQSITLTLESPPGSSPSVQCRSPRKNIO----- 164
Db      2135 VOLNVQRAAANARITGTSPPRTDVRYGGLTKLDCSASGDPRIILMLPSKRMIDALFSF 2194
Qy      165 -----GGKTLVSQLELDQSGTWCTVLONQ---KVEFKIDIYVLAFQKASSIYKK 214
Db      2195 DSRIRKVPANGTLVVKSVTDKADGDLIC-VARNKVGDYVVLKVDVVM---KPAKIEHKE 2249
Qy      215 EGE-----QVERFPLATFYEKLTGSGELMWQABRASSSKSMIT 253
Db      2250 ENDRKVFYGGDLKVDCAVATGLPNPEISWSLPDGLVNSFWQSD-----DSGGRTKRYV 2303
Qy      254 FD---LKNKEVSVKRVTDPKLOMGKKLPLHLTLPOALPOYAGSGNLTLEAKTGKHLQ 310
Db      2304 FNNGTLYFNEVGME-----EGDYTCFAENQVGDDEM 2335
Qy      311 EVNLVVRATOLQKNLT-----CEVWGPTSPKMLSLKLENKEAKSVKRE 355
Db      2336 RVRKVVTAPATIRNKTYLAQVPEYGDVTVVACEAKGEMPKVTW-LSPTNKVIPTSSSEK 2394
Qy      356 KPVW-----VLNPEAGMWOCILSDSG---QVLESNIKVLPTWSPVHPRASAL 401
Db      2395 YQIYQDGTLLIQKQKSDSGNYTCLVRNSAGEDRKTYWHVNVQ----- 2438
Qy      402 PAPPTGSALPDPOT-----ASALPDPPEASALPALAV 434
Db      2439 --PPKINGNPNPITTVREIAAGSRKLLDCKAEGIPFPVLMAPPEGVVL 2486

```

```

RESULT 88
US-10-025-027-58
; Sequence 58, Application US/10295027

```



```

1 Publication No. US20030232350A1
2 GENERAL INFORMATION:
3 APPLICANT: Afar, Daniel
4 APPLICANT: Aziz, Natasha
5 APPLICANT: Gineberg, Wendy M.
6 APPLICANT: Gish, Kurt C.
7 APPLICANT: Glynn, Richard
8 APPLICANT: Hevezl, Peter A.
9 APPLICANT: Mack, David H.
10 APPLICANT: Murray, Richard
11 APPLICANT: Watson, Susan R.
12 APPLICANT: Eos Biotechnology, Inc.
13 TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
14 FILE OF INVENTION: Methods of Screening for Modulators of Cancer
15 FILE REFERENCE: 018501-012500US
16 CURRENT APPLICATION NUMBER: US/10/295,027
17 CURRENT FILING DATE: 2002-11-13
18 PRIOR APPLICATION NUMBER: US 09/663,733
19 PRIOR FILING DATE: 2000-09-15
20 PRIOR APPLICATION NUMBER: US 60/350,666
21 PRIOR FILING DATE: 2001-11-13
22 PRIOR APPLICATION NUMBER: US 60/335,394
23 PRIOR FILING DATE: 2001-11-15
24 PRIOR APPLICATION NUMBER: US 60/332,464
25 PRIOR FILING DATE: 2001-11-21
26 PRIOR APPLICATION NUMBER: US 60/334,393
27 PRIOR FILING DATE: 2001-11-29
28 PRIOR APPLICATION NUMBER: US 60/340,376
29 PRIOR FILING DATE: 2001-12-14
30 PRIOR APPLICATION NUMBER: US 60/347,211
31 PRIOR FILING DATE: 2002-01-08
32 PRIOR APPLICATION NUMBER: US 60/347,349
33 PRIOR FILING DATE: 2002-01-10
34 PRIOR APPLICATION NUMBER: US 60/355,250
35 PRIOR FILING DATE: 2002-02-08
36 PRIOR APPLICATION NUMBER: US 60/356,714
37 PRIOR FILING DATE: 2002-02-13
38 Remaining Prior Application data removed - See file wrapper or PALM.
39 NUMBER OF SEQ ID NOS: 1386
40 SOFTWARE: PatentIn Ver. 2.1
41 SEQ ID NO: 58
42 LENGTH: 2828
43 TYPE: PRT
44 ORGANISM: Homo sapiens
45 US-10-295-027-58
46
47 Query Match 5.9%; Score 138; DB 15; Length 2828;
48 Best Local Similarity 19.4%; Pred. No. 0.5;
49 Matches 103; Conservative 63; Mismatches 180; Indels 184; Gaps 21.
50
51 10 LLLVQLALLPAAATGKNV--VLKGGGTVELTCTASQKKSIQFMKNSNQIKLGGNGS 67
52 2036 LAIRLHVALLPVLVHQEKLENISLPPGISIRHCTAKAAPLPSVW-----VLG----- 2084
53
54 68 FLTKGPSKLNDRAISRSLMDGN---FP---LIINKLIEDSDTYICE---VEDQKEE 117
55 2085 -----DOTGRPSQFLHGNI FVPENGTLYIRNLAPKDSGRYECVAANLVGSART 2133
56
57 118 VOLLVFGLTANS-----DTHLQGSLLTTLTSPGSSPSVQCRSPRGKNI----- 164
58 2135 VOLVQRAAANAARITGTSPPRRIDVYGGTLKIDCSASGDPWPRILMRPLPSKMDIALPSF 2194
59
60 165 -----GKTLTSVSGLELDGSGTCTCVLQNO---KVEFIDIVLVAFOKASIVYKK 214
61 2195 DSRIVFPANGILVKSVDIKDGDYLC-VARKKVGDDYVVLKVDVW-----KPAITEKE 2249
62
63 215 EGE-----QVEFSFLAFTVEKLTGSGELMWQAEARASSSKSWIT 253
64 2250 ENDHKVFPYGGDLKVDCAVATGLPNEPISWSLPPGSLVNSFWQSD-----DSGRTRKRVV 2303
65
66 254 FD---LKRKEYSVKRVTQDPKLQMGKKLPLHLTLFQALPYAGSGNLTLLAEAKTGKJHQ 310
67 2304 FNNGTLYFNEVGMRB-----EDDYTCFAPENGVGKDEM 2335

```

```

Qy      311  EVNLVWMEATQLQKXLT-----CEWVGPTSKMLSLTLENKAAYSKRE 355
Db      2336  RRRVRAVPAATIRIKTKTLVAQVPYGVAVTYACEAKGPMPEKWTW-LSPTNVVIPTSSK 2394
Qy      356  KPWW-----VLNPEAGMWCILSDSG-----QVLLSBNIKVLPWMTSPVHPRASAL 401
Db      2395  YQIVDDGTLLIQKQNRSDSGNYTCLVRRSAGEDRTVIAHHNVQ----- 2438
Qy      402  PAPPTGSLPDPQT-----ASALDPPEASALPALAV 434
Db      2439  --PFXKNGNPNPITTVREIRIAGSGSRKLIDCKAEIGTPRVLVMAFEEGVVL 2486

RESULT 89
US-10-295-027-1175
/ Sequence 1175, Application US/10295027
/ Publication No. US20030232350A1
/ GENERAL INFORMATION:
/ APPLICANT: Afar, Daniel
/ APPLICANT: Aziz, Natasha
/ APPLICANT: Ginsberg, Wendy M.
/ APPLICANT: Gish, Kurt C.
/ APPLICANT: Glynn, Richard
/ APPLICANT: Hevezi, Peter A.
/ APPLICANT: Mack, David H.
/ APPLICANT: Murray, Richard
/ APPLICANT: Watson, Susan R.
/ APPLICANT: Eos Biotechnology, Inc.
/ TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
/ FILE REFERENCE: 018501-012500US
/ CURRENT APPLICATION NUMBER: US/10/295,027
/ CURRENT FILING DATE: 2002-11-13
/ PRIOR APPLICATION NUMBER: US 09/663,733
/ PRIOR FILING DATE: 2000-09-15
/ PRIOR APPLICATION NUMBER: US 60/350,666
/ PRIOR FILING DATE: 2001-11-13
/ PRIOR APPLICATION NUMBER: US 60/335,394
/ PRIOR FILING DATE: 2001-11-15
/ PRIOR APPLICATION NUMBER: US 60/332,464
/ PRIOR FILING DATE: 2001-11-21
/ PRIOR APPLICATION NUMBER: US 60/334,393
/ PRIOR FILING DATE: 2001-11-29
/ PRIOR APPLICATION NUMBER: US 60/340,376
/ PRIOR FILING DATE: 2001-12-14
/ PRIOR APPLICATION NUMBER: US 60/347,211
/ PRIOR FILING DATE: 2002-01-08
/ PRIOR APPLICATION NUMBER: US 60/347,349
/ PRIOR FILING DATE: 2002-01-10
/ PRIOR APPLICATION NUMBER: US 60/355,250
/ PRIOR FILING DATE: 2002-02-08
/ PRIOR APPLICATION NUMBER: US 60/356,714
/ PRIOR FILING DATE: 2002-02-13
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 1386
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1175
/ LENGTH: 2828
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-295-027-1175

Query Match      5.9%; Score 138; DB 15; Length 2828;
Best Local Similarity 19.4%; Pred. No. 0.5;
Matches 103; Conservative 63; Mismatches 180; Indels 184; Gaps 21;

Qy      10  LLVLVGLALPLPAATQGNKV-VLKKGGTVELTCTASQSKSIQPHMKNSNGIKILGNGS 67
Db      2036  LAIRLHVAAALPPVHQEKLENIISLPGSLIIRHCTAKAAPLPSPRW-----VLG----- 2084
Qy      68  FLTKGPKSLANDRADSRRLMDQGN---FP---LIIRKLIKEDSDTYICE---VEDOKEE 117

```

```

Db 2085 -----DGTQIRPSQFLHGNLFVFPNGTLYIRNLAPRSGRYECVAAANLVGSART 2134
Qy 118 VQLLVFGLTANS-----DTHLLQGSLTFLTESPPGSSPVQCRSPRKNIG----- 164
Db 2135 VQLNVQRAAANARITGTSPPRTDVRIGGTLKIDCSASGDPFRLMLRPSKRMIDALPSF 2194
Qy 165 -----GKTLVSQLELQDSGTWCTVLQNO--KKEPEKIDIVLAFQKASSIYKK 214
Db 2195 DSRKIVFANGTLVVKSVTDKQAGDYLQ-VARNKVGDDIVVLKVDVVM---KPAKIBHKE 2249
Qy 215 EGE-----QVEFSPLAFTVEKLTSSGELMWAQEAASSKSKMIT 253
Db 2250 ENDKHVFYGGDLKVDCAVATGLPNPEISWSLDPGSLVNSFMQSD-----DSGGRTKAYV 2303
Qy 254 FD---LNKKEYSVKRVTDQDFLQMGKKLPLHLTLPLQALPQVAGSGNLTALFAKTKLHQ 310
Db 2304 FNNGLTYNEVGMB-----EGDTTCFANOVGDDEM 2335
Qy 311 EVNLVVRATQLQKRLT-----CEVWGPTSPKMLSLKENKEAKVSKRE 355
Db 2336 RVRVKVVTAPATIRNKTYLAVQVPYGVVTVACAKGKPMKVTW-LSPTKKVIPTSSEK 2394
Qy 356 KPVW-----VLNPRAGMWQCLISDSG-----QVLLSENIIVLPWSTPHVPRASAL 401
Db 2395 YQIYQDGTLLIQKQKQSDSGNYTCLVNSAGEDRKTYWIHVNO----- 2438
Qy 402 PAPPGSALPDPQF-----ASALPDPASALPALAV 434
Db 2439 --PKKINGNPVITTVREIAGSGRKLIDCAEGIPTRPVLMARPEGVVL 2486

```

```

RESULT 90
US-10-253-286-282
; Sequence 282, Application US/10253286
; Publication No. US20040058881A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: II-KEY/ANTIGENIC EPIOTOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REF-2015
; CURRENT APPLICATION NUMBER: US/10/253,286
; CURRENT FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 282
; LENGTH: 702
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-253-286-282

```

```

Query Match 5.9%; Score 137.5; DB 12; Length 702;
Best Local Similarity 18.8%; Pred. No. 0.081;
Matches 96; Conservative 76; Mismatches 215; Indels 123; Gaps 20;

```

```

Qy 20 PAATQGNKVVLAGKGGDTVELTCTASQKKSIOFPMKNSNOIKILNGQSFITKGPSKLNDR 79
Db 147 PSISSNSKPEDEK-DAVAFTCEPETQDATYLMVWVNOQLPV----- 187
Qy 80 ADSRSLMDQGNFPLIKLKIEDSDTYICEVED-----QKEVQL-LVFGLTANS----- 129
Db 188 --SRLQLSNGKRLTLFLNVRANDTASCKCETQNPVARSRSVILNVLGPDPATISPL 245
Qy 130 DTHLLQGSLTFLTESPPGSSPVQCR-SPRKNIGGKTLVSQLELQDSGTWCTVLQ 188
Db 246 NTSYRSGENLNLSCHA--ASNPPAQYSWFGTFQOSTQELFIPNITVNNSGSYTQCAHN 303
Qy 189 NQKV-EFKIDIVLAFQKASSIYKKEGEVFSPLAFTVE-KLTGSGELMWQABAS 246
Db 304 SDTGLNRTTITTVYAEPPKPFITSNNSNVEDEDAVALTCPEIQNTTYLWVWVNOQL 363

```

```

Qy 247 SSKSWITFDLKNKEVSVKRVTDQ-----KLQMGKKLPLHL-----TLPOA 287
Db 364 PVSPRLQLSNDKRLTLFLSTVRNDVGPYEGGIGNELSVHSDVILNVLXGPDPTISPS 423
Qy 288 LPQYAGSGNLTALFAKT-----GKLHQEV-NLVVRATQLQKRL-TCRWGPT 334
Db 424 YTYRRGVNLISLSCHAASNPQAYSWLIDGNIQHQHQLFISNITEKNSGLYTCQANNSA 483
Qy 335 SPKMLSLKENKEAV-----SKREKPV-----WVLNPRAGMWQCLISSGQVTLLE 381
Db 484 SGHSRTVTKITTVYSAELPKPSISSNSKVEDKDAVAFTCEPEAQNTTYLWVWVNOQ---- 539
Qy 382 SNIKVLPTWSTPHVPRASALPAPT-----GSALPDPOTASAL 419
Db 540 -----SLPSPRLQLSNGKRLTLFLNVRANDARAYVCGIONSARSRDPVTLVDL 530
Qy 420 --DPPASALPALAVISFLGLGVAC 447
Db 591 YGPDTPILSPDS-----SYLSGANLNLSC 615

```

```

RESULT 91
US-10-380-136-16
; Sequence 16, Application US/10380136
; Publication No. US20040071726A1
; GENERAL INFORMATION:
; APPLICANT: Chiciz, Roman M.
; APPLICANT: Tomlinson, Andy
; TITLE OF INVENTION: PEPTIDE EPIOTOPES RECOGNIZED BY ANTIGEN
; FILE REFERENCE: 08191-019US1
; CURRENT APPLICATION NUMBER: US/10/380,136
; CURRENT FILING DATE: 2003-03-11
; PRIOR APPLICATION NUMBER: PCT/US01/28467
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: US 60/232,185
; PRIOR FILING DATE: 2000-09-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 702
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-380-136-16

```

```

Query Match 5.9%; Score 137.5; DB 12; Length 702;
Best Local Similarity 18.8%; Pred. No. 0.081;
Matches 96; Conservative 76; Mismatches 215; Indels 123; Gaps 20;

```

```

Qy 20 PAATQGNKVVLAGKGGDTVELTCTASQKKSIOFPMKNSNOIKILNGQSFITKGPSKLNDR 79
Db 147 PSISSNSKPEVEK-DAVAFTCEPETQDATYLMVWVNOQLPV----- 187
Qy 80 ADSRSLMDQGNFPLIKLKIEDSDTYICEVED-----QKEVQL-LVFGLTANS----- 129
Db 188 --SRLQLSNGKRLTLFLNVRANDTASCKCETQNPVARSRSVILNVLXGPDPTISPL 245
Qy 130 DTHLLQGSLTFLTESPPGSSPVQCR-SPRKNIGGKTLVSQLELQDSGTWCTVLQ 188
Db 246 NTSYRSGENLNLSCHA--ASNPPAQYSWFGTFQOSTQELFIPNITVNNSGSYTQCAHN 303
Qy 189 NQKV-EFKIDIVLAFQKASSIYKKEGEVFSPLAFTVE-KLTGSGELMWQABAS 246
Db 304 SDTGLNRTTITTVYAEPPKPFITSNNSNVEDEDAVALTCPEIQNTTYLWVWVNOQL 363
Qy 247 SSKSWITFDLKNKEVSVKRVTDQ-----KLQMGKKLPLHL-----TLPOA 287
Db 364 PVSPRLQLSNDKRLTLFLSTVRNDVGPYEGGIGNELSVHSDVILNVLXGPDPTISPS 423
Qy 288 LPQYAGSGNLTALFAKT-----GKLHQEV-NLVVRATQLQKRL-TCRWGPT 334
Db 424 YTYRRGVNLISLSCHAASNPQAYSWLIDGNIQHQHQLFISNITEKNSGLYTCQANNSA 483

```

```
QY 335 SPKMLSLKLENKEAKV-----SKREKPV-----WVLANPEAGMOCILSDSGOVLLE 381
DB 484 SGHSRTTYKTTIVSALEPKPSSISNNKRVEDKDAVAFCEPAQNTTYLWMVNGQ----- 539
QY 382 SNIKVLPWTSTVPHRASALPAPPT-----GSALEPDPTASAL 419
DB 540 -----SLPVSRLQLSNGNRITLTLFNTVRNDARAAYVCGIQNSVANSRSDPVTLDVL 590
QY 420 --PDPPAASALPALAVISFLLGLGLGVAC 447
DB 591 YGPDPTIISPDS-----SYLSGANLNLISC 615

RESULT 92
US-10-157-031-341
; Sequence 341, Application US/10157031
; Publication No. US20030108890A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Yankovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobachev, A. V.
; APPLICANT: Krukovekaya, L. L.
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 341
; LENGTH: 702
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-157-031-341

Query Match          5.9%; Score 137.5; DB 14; Length 702;
Best Local Similarity 18.8%; Pred. No. 0.081; Indels 123; Gaps 20;
Matches 96; Conservative 76; Mismatches 215;

QY 20 PAATQGNKRVLGKKDVELTCTASQKSIQPHWKNNOIKILGNQGSFLTGGPSKLNDR 79
DB 147 PSISNNKRPVEDK-DAVAFCEPETQDATYLMWVNGSLPV----- 187
QY 80 ADSRSLMDQGNFPLIINKLKIEDSDTYICEVED-----QKEEVQL-LVFGLTANS--- 129
DB 188 --SPRLQLSNGNRITLTLFNTVRNDTASYKCEQTQNPVSARRSDSVILNLVYGPDAPTISPL 245
QY 130 DTHLLQGSILTLTLESPGSSPSVQCR-SPRGKNIQGGKTLVSQLELDGSGTWCTVYLQ 188
DB 246 NTSYRSGENLNLSCHA--ASNPPAQYSFWNGTFOQSTQELFPNITVNNSSGYTCQAHN 303
QY 189 NQKV-EFKIDIVLAFQKASSIVYKKEGEQVEFSPLAFVE-KLTGSGELMWQAERAS 246
DB 304 SDTGNRRTVTITTYAERPKPFITSNNSNPVEDDVAALICEPEIQNTTYLWMVNGSL 363
QY 247 SSKSWITDCLKKEVSVKRVTDPP-----KLGKGLPLHL-----TLPOA 287
DB 364 FVSPRLQLSNDNRITLTLFNTVRNDVGPYECGIONELSVDSHPVILNLVYGPDDPTISPS 423
QY 288 LPOYAGSGLTLALAEKT-----GKLHQEV-NLVVMRATOLQKNL-TCEVMGPT 334
DB 424 YTYRPGVNLSLCHASNPRAQYSMLIDGNIQHTQELFISNITEKNSGLTYCQANNSA 483
QY 335 SPKMLSLKLENKEAKV-----SKREKPV-----WVLANPEAGMOCILSDSGOVLLE 381
DB 484 SGHSRTTYKTTIVSALEPKPSSISNNKRVEDKDAVAFCEPAQNTTYLWMVNGQ----- 539
QY 382 SNIKVLPWTSTVPHRASALPAPPT-----GSALEPDPTASAL 419
DB 540 -----SLPVSRLQLSNGNRITLTLFNTVRNDARAAYVCGIQNSVANSRSDPVTLDVL 590
QY 420 --PDPPAASALPALAVISFLLGLGLGVAC 447
```

```
DB 591 YGPDPTIISPDS-----SYLSGANLNLISC 615

RESULT 93
US-10-207-655-87
; Sequence 87, Application US/10207655
; Publication No. US20030118592A1
; GENERAL INFORMATION:
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Hayden-Ledbetter, Martha S.
; TITLE OF INVENTION: BINDING DOMAIN-IMMUNOGLOBULIN FUSION PROTEINS
; FILE REFERENCE: 390069.401C1
; CURRENT APPLICATION NUMBER: US/10/207,655
; CURRENT FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 426
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 87
; LENGTH: 702
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-207-655-87

Query Match          5.9%; Score 137.5; DB 14; Length 702;
Best Local Similarity 18.8%; Pred. No. 0.081; Indels 123; Gaps 20;
Matches 96; Conservative 76; Mismatches 215;

QY 20 PAATQGNKRVLGKKDVELTCTASQKSIQPHWKNNOIKILGNQGSFLTGGPSKLNDR 79
DB 147 PSISNNKRPVEDK-DAVAFCEPETQDATYLMWVNGSLPV----- 187
QY 80 ADSRSLMDQGNFPLIINKLKIEDSDTYICEVED-----QKEEVQL-LVFGLTANS--- 129
DB 188 --SPRLQLSNGNRITLTLFNTVRNDTASYKCEQTQNPVSARRSDSVILNLVYGPDAPTISPL 245
QY 130 DTHLLQGSILTLTLESPGSSPSVQCR-SPRGKNIQGGKTLVSQLELDGSGTWCTVYLQ 188
DB 246 NTSYRSGENLNLSCHA--ASNPPAQYSFWNGTFOQSTQELFPNITVNNSSGYTCQAHN 303
QY 189 NQKV-EFKIDIVLAFQKASSIVYKKEGEQVEFSPLAFVE-KLTGSGELMWQAERAS 246
DB 304 SDTGNRRTVTITTYAERPKPFITSNNSNPVEDDVAALICEPEIQNTTYLWMVNGSL 363
QY 247 SSKSWITDCLKKEVSVKRVTDPP-----KLGKGLPLHL-----TLPOA 287
DB 364 FVSPRLQLSNDNRITLTLFNTVRNDVGPYECGIONELSVDSHPVILNLVYGPDDPTISPS 423
QY 288 LPOYAGSGLTLALAEKT-----GKLHQEV-NLVVMRATOLQKNL-TCEVMGPT 334
DB 424 YTYRPGVNLSLCHASNPRAQYSMLIDGNIQHTQELFISNITEKNSGLTYCQANNSA 483
QY 335 SPKMLSLKLENKEAKV-----SKREKPV-----WVLANPEAGMOCILSDSGOVLLE 381
DB 484 SGHSRTTYKTTIVSALEPKPSSISNNKRVEDKDAVAFCEPAQNTTYLWMVNGQ----- 539
QY 382 SNIKVLPWTSTVPHRASALPAPPT-----GSALEPDPTASAL 419
DB 540 -----SLPVSRLQLSNGNRITLTLFNTVRNDARAAYVCGIQNSVANSRSDPVTLDVL 590
QY 420 --PDPPAASALPALAVISFLLGLGLGVAC 447
DB 591 YGPDPTIISPDS-----SYLSGANLNLISC 615

RESULT 94
US-10-117-937-592
; Sequence 592, Application US/10117937
; Publication No. US20030220239A1
; GENERAL INFORMATION:
; APPLICANT: CTL IMMUNO THERAPIES CORP.
; APPLICANT: SIMARD, John, J. L.
; APPLICANT: DIAMOND, David, C.
; APPLICANT: Liu, Liping
```



```
Db          591 YGDPTRPISPPDS-----SYLSGANLNLSC 615

RESULT 96
US-10-245-871-282
; Sequence 282, Application US/10245871
; Publication No. US20030235594A1
; GENERAL INFORMATION:
; APPLICANT: HUMPHREYS, ROBERT
; APPLICANT: XU, MINZHEN
; TITLE OF INVENTION: I1-KEY/ANTIGENIC EPI TOPE HYBRID PEPTIDE VACCINES
; FILE REFERENCE: REH-2013
; CURRENT APPLICATION NUMBER: US/10/245,871
; PRIOR APPLICATION NUMBER: 10/197,000
; PRIOR FILING DATE: 2002-07-17
; PRIOR APPLICATION NUMBER: 09/396,813
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 905
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 282
; LENGTH: 702
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-245-871-282

Query Match          5.9%; Score 137.5; DB 15; Length 702;
Best Local Similarity 18.8%; Pred. No. 0.081;
Matches 96; Conservative 76; Mismatches 215; Indels 123; Gaps 20;

Oy 20 PAATQGNKVVILGKGDVTELCTASQKSIQFHKNSNQIKILNQSGFLTKGSKLND 79
Db 147 PSISSNNSKPEVDK-DAVAFTCEPETO DATYILMWNQSLPV----- 187
Oy 80 ADSRRSLMDQGNFPLIILKLIKIEDSDTYICEVED-----QKEEYOL-LVFGLTANS 129
Db 188 --SPRLQSLNSGNRTLTLEFNVRNDTASVKCEIQNVASRRSDSYLVNLVYGPDPATISPL 245
Oy 130 DTHLLQSGSLTTLTLESPGSSPSVQCR-SPRGKNIQGGKTLVSQLELQDSGTWTCVTLQ 188
Db 246 NTSYRSGENLNLSCHA--ASNPPAQYSWFVNGTFQSGSTOELEFIPNITVNNSGSYTQAHN 303
Oy 189 NQKKV-EFKIDIVLAFQKASSIVYKKEGEVSEFPLAFYVE-KLTGSGELMWQAERAS 246
Db 304 SDTGILNRTVTITTVYAEPPKPFITSNNSNPFVEDDAVALTCEPEIQNTYILMWNQSL 363
Oy 247 SSKSWITFDLKNKEVSVKRVTDQ-----KLGQKGLPLHL-----TLPOA 287
Db 364 PVSPPRLQSLNSGNRTLTLEFNVRNDVGYECGIQNELSVDSHPVILNLVYGPDDPTISPS 423
Oy 288 LPOVAGSGNLTLALFAKT-----GKLHQEV-NLVVMRATOLQKNL-TCEVWGPT 334
Db 424 YTYVYRPGVNLISCHASNPAPQYQWMLIDGNIQHTQELFISNITEKNSGLYTQANNSA 483
Oy 335 SPKMLSLKLEKKEAKV-----SKREKPY-----WLNPEAGMOCCLSDSQVULE 381
Db 484 SGHSRTTKTITVSAELPKPSISSNNSKPEVDKDAVAFTCEPEAQNTTYILMWNQSL 539
Oy 382 SNIKVLPWSTFVHRRASALPAPPT-----GSALEPDPOTASAL 419
Db 540 -----SLPVSRLQSLNSGNRTLTLEFNVRNDARAYVCGIQNSVANSNDPVTLDVL 590
Oy 420 --PDPPAASALPALALAVISFLIGLGVAC 447
Db 591 YGDPTRPISPPDS-----SYLSGANLNLSC 615

RESULT 97
US-09-756-551A-17
; Sequence 17, Application US/09756551A
; Patent No. US20020051768A1
; GENERAL INFORMATION:
; APPLICANT: C. Morrow et al.
; TITLE OF INVENTION: ENCAPSIDATED RECOMBINANT VIRAL
; TITLE OF INVENTION: NUCLEIC ACID AND METHODS OF MAKING AND
; TITLE OF INVENTION: USING SAME
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/756,551A
; FILING DATE: 08-JAN-2001
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/376,184
; FILING DATE: 17-AUG-1999
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/987,867
; FILING DATE: 09-DEC-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/389,459
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/087,009
; FILING DATE: 01-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauro, Peter C.
; REGISTRATION NUMBER: 32,360
; REFERENCE/DOCKET NUMBER: UAI-004CPDVZCN
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 734 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-756-551A-17

Query Match          5.9%; Score 137.5; DB 9; Length 734;
Best Local Similarity 18.8%; Pred. No. 0.086;
Matches 96; Conservative 76; Mismatches 215; Indels 123; Gaps 20;

Oy 20 PAATQGNKVVILGKGDVTELCTASQKSIQFHKNSNQIKILNQSGFLTKGSKLND 79
Db 179 PSISSNNSKPEVDK-DAVAFTCEPETO DATYILMWNQSLPV----- 219
Oy 80 ADSRRSLMDQGNFPLIILKLIKIEDSDTYICEVED-----QKEEYOL-LVFGLTANS 129
Db 220 --SPRLQSLNSGNRTLTLEFNVRNDTASVKCEIQNVASRRSDSYLVNLVYGPDPATISPL 277
Oy 130 DTHLLQSGSLTTLTLESPGSSPSVQCR-SPRGKNIQGGKTLVSQLELQDSGTWTCVTLQ 188
Db 278 NTSYRSGENLNLSCHA--ASNPPAQYSWFVNGTFQSGSTOELEFIPNITVNNSGSYTQAHN 335
Oy 189 NQKKV-EFKIDIVLAFQKASSIVYKKEGEVSEFPLAFYVE-KLTGSGELMWQAERAS 246
Db 336 SDTGILNRTVTITTVYAEPPKPFITSNNSNPFVEDDAVALTCEPEIQNTYILMWNQSL 395
Oy 247 SSKSWITFDLKNKEVSVKRVTDQ-----KLGQKGLPLHL-----TLPOA 287
Db 396 PVSPPRLQSLNSGNRTLTLEFNVRNDVGYECGIQNELSVDSHPVILNLVYGPDDPTISPS 455
```

```

Qy 288 LPOYAGSGLTLALEAKT-----GKLEHOEV-NLVVMRATOLQKNL-TCEVWGPT 334
Db 456 YTYRPGVNLISCHMAASNPQAQYKWLIDGNIQCHTOGLFSNTERKSSGLYTQANNSA 515
Qy 335 SPKMLSLKENKEAKV-----SKREKPV-----WLNPEAGMOCCLISDSGOVLE 381
Db 516 SGHSRTTVKXTITVAEELPKPSISSNNSKPEVDKDAVAFCEPAQNTTYLWMVANGQ---- 571
Qy 382 SNIKVLPWSTPVHPRASALPAPT-----GSAALPDQTASAL 419
Db 572 -----SLPVSERLOLSNGNRTLTLEFNTVRDARAYVCIGIONSVANSRSDVTLTDLV 622
Qy 420 --PDPPASALPALAVISFLGLGLGVAC 447
Db 623 YGPDPTIISPPDS-----SYLSGANLNLSC 647

RESULT 98
US-10-451-168-78
; Sequence 78, Application US/10451168
; Publication No. US20040091969A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: SMITHKLINE BEECHAM P.L.C.
; APPLICANT: GLAXO GROUP LIMITED
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50039
; CURRENT APPLICATION NUMBER: US/10/451,168
; CURRENT FILING DATE: 2003-11-12
; PRIOR APPLICATION NUMBER: PCT/US01/49232
; PRIOR FILING DATE: 2000-12-17
; PRIOR APPLICATION NUMBER: 60/256,710
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/257,048
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/260,482
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/264,922
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/266,797
; PRIOR FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: 60/276,988
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/281,535
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/289,622
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 5635
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-451-168-78

Query Match 5.9%; Score 137; DB 16; Length 5635;
Best Local Similarity 20.4%; Pred. No. 1.5;
Matches 93; Conservative 54; Mismatches 170; Indels 138; Gaps 21;

```

```

Db 964 NVAGTNNKTSVAVVHLPRIHQGQILSTIEG--IPVTLPE-----CKA 1004
Qy 245 ASSSKSWITDNLKNKEVSRYRVQDPKQMGKLPPLHLTPOLAPYAGSGLTLALEAK 304
Db 1005 SGNPKBSVLMWSKKGELIS-----TSSAKFSAAGDSLYVSPGG---EESGEYVCTATNT 1056
Qy 305 TGKLEHOEVNLV-----MRATOLQK-----NLTCVEV-----W--- 331
Db 1057 AGYAKKRVQGLTVVRRVFDQGLSQDKPVEISVLAGEVTLPCFEYKSLPPLITWAKE 1116
Qy 332 ----GPTSPK--LMLSILKENKEAKVSKREKPVVNLNPEAGMOCCLISD--SGOV--LLES 382
Db 1117 TQIISPSRPHRTFLPGSGMKITETRTS-----DSGMVLCAVTNAGVNTQAVKL 1165
Qy 383 NIKVLPWST--PVHPRASA-----LPAPPGSALP 411
Db 1166 NVHVPKIQRGPRHLKRVGQRVDPICNAGTLP 1200

RESULT 99
US-10-106-698-4871
; Sequence 4871, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 4871
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (392)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (393)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC FEATURE
; LOCATION: (571)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-4871

Query Match 5.9%; Score 136.5; DB 14; Length 595;
Best Local Similarity 18.5%; Pred. No. 0.077;
Matches 95; Conservative 75; Mismatches 214; Indels 129; Gaps 20;

```

```

OY      247 SSKSMITFDLKKKKKSVKRVNODP-----KIQMGKKKPLHL-----TLPGA 287
Db      257 PVSPPRLQSLNDNRKRTLLSLVTRNDVGPYEGCIGIONELSVDSHSDVILNLVYGPDDPTTSPS 316
OY      288 LPQVAGSGNLTLLALEAKT-----GKLHQEV-NLVVMRATOLQKLT-TCSEWGP 334
Db      317 YTYVYRPGVNLISLSCHAAKNPAQVSWLIDGNIQHQHQLFISNITEKNSGLTCQANNSA 376
OY      335 S-----PKLMLSLKLENKAEKYSKREKPVWLNPEAGMKQCLLSGGV 378
Db      377 SGHSRTTYKTTIVSAXXPKPSIS--SNNSKPEVEDKVAIFCEPBAQNTTYLIMWVNGQ- 432
OY      379 LLESNIKVLPTWSTPVHPRASALPAEPT-----GSAALPDPOZA 416
Db      433 -----SLVPSRPLQSLNSNGRRTLLFNVTNRNDARAYVCGIONSANRSPPVTL 480
OY      417 SAL--PDPPAALPALPALANVISFLGLGIGVAC 447
Db      481 DVLVGPDPPTTIIISPDS---SYLSGANINLISC 508

```

```

RESULT 100
: US-09-808-602-67
: Sequence 67, Application US/09808602
: Patent No. US2002015115A1
: GENERAL INFORMATION:
: APPLICANT: Vernet, Corine A
: APPLICANT: Fernandes, Elma
: APPLICANT: Shinkets, Richard A
: APPLICANT: Herrman, John L
: APPLICANT: Majumder, Kumud
: APPLICANT: Mishra, Vishnu
: APPLICANT: Mezes, Peter S
: APPLICANT: MacDougall, John
: TITLE OF INVENTION: No. US2002015115A1el Proteins and Nucleic Acids Encoding Same
: FILE REFERENCE: 15966-697 CIP
: CURRENT APPLICATION NUMBER: US/09/808, 602
: CURRENT FILING DATE: 2001-03-14
: PRIOR APPLICATION NUMBER: 09/800, 198
: PRIOR FILING DATE: 2001-03-05
: PRIOR APPLICATION NUMBER: 60/146, 596
: PRIOR FILING DATE: 2000-03-03
: NUMBER OF SEQ ID NOS: 114
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 67
: LENGTH: 1395
: TYPE: PRT
: ORGANISM: Drosophila melanogaster
: US-09-808-602-67

```

Query Match	5.9%;	Score 136.5;	DB 9;	Length 1395;
Best Local Similarity	21.1%;	Pred. No. 0.25;		
Matches 102;	Conservative 53;	Mismatches 179;	Indels 149;	Gaps 18

```

QY 26 KKVVLGKKGDPTVELTCTSAQKKSIOQHMKNSNOIKILNGQSGLTGKSPKLNDRADRS 85
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 NKV-GLNG-VQLEPCMASGNPPSYFTWK-----EGVSTLWFPNSHGR 395
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 86 LMDQGNFLLIKNLKIEDSTYICEVEDQKEVQLLVFGLTANSPTN---LLOGSGLTTLT 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 396 QYVADGTGLQITDIVEOEDEGYVVCASFVSVDSTYVFLQVSSVDPERPPIIDIGANOT 455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 143 LESPPGSSPSVOCR-----SPRGK-----NIOGKTLVSQLELDQSDGTWT 183
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 456 L-FGGSVATLEPCRATGNPSPRIKMWFHDHAVAQGNRYSIIGSSLRVDDLQSDSGTYT 513
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 184 CTVLONQKVEKIKIDIVLAFQKASIVYKKGGEVPEPPLAFTEKLTGSGELWQME 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 514 CTA-----SEERGETSWAATLIVLEK-PGDSISLRAD 544
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 244 -----RASSSKSWITFDLKNKEVS-VKRVTD-----PKLQMGKPLH 281
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db      545 PSTYTPAPGPKLVANSTRISILRNAAKSGEKGAVPIIGTVAYFSPDQLQTGWIVAAHR 604
      282 -----LTLFQALPY-----AGSGNTTLA 303
Qy      605 VEDTQVTTISGLTPGTSYVELVRAENTOGISVPSGLSNVIKTI EADFPDAASANDLSAARTL 664
      304 KTGKLEHGVNLVVMRATOLOKNLTCEWNGPSPKLMLSLKLENKEAKYSKREKEPVWLNP 363
Qy      665 LTGKSVELLDAISAINSAVRLMWMLHV--SADEKTVESGLRIHYHDASY-----P 711
Db      364 EAGMOCCLSD--SGQVLTESNIKVLPTWSTVPHRASALPAPPTGSALPDPQTASALPDP 422
Qy      712 SAQYHSITVMDASAESEFVGNLKKTTKVEFLPTTFETIEGQPSNS-----KTALYEDV 766
Db      423 PAA 425
      767 PSA 769

```

QY	423	PAA	425
Db	767	PSA	769

Search completed: August 3, 2004, 13:47:44  
Job time : 45.3425 secs

**THIS PAGE BLANK (USPTO)**





103	282.5	11.7	348	2	S09273	Ig alpha chain C r
104	282	11.7	347	2	S09274	Ig alpha chain C r
105	281	11.6	353	1	HVRKc1	Ig mu chain C regi
106	280.5	11.6	99	2	S521462	T-cell surface gly
107	279	11.6	352	2	S055500	Ig alpha-1 chain C
108	277.5	11.5	353	1	A1Hu	Ig alpha-1 chain C
109	277	11.5	340	1	A2Hu	Ig alpha-2 chain C
110	275.5	11.4	88	2	A30503	Ig gamma-2b chain
111	275	11.4	357	2	S09267	Ig alpha chain C r
112	275	11.4	568	2	A45804	Ig mu chain C regi
113	268.5	11.1	350	2	S09271	Ig alpha chain C r
114	264.5	11.0	344	1	AHMS	Ig alpha chain C r
115	263.5	10.9	358	2	S09268	Ig alpha chain C r
116	258	10.7	580	2	A46538	Ig heavy chain C r
117	257.5	10.6	342	2	A45966	Ig heavy chain, se
118	257	10.6	111	2	A43148	Ig alpha chain C r
119	253.5	10.5	573	2	S12838	Ig mu chain precu
120	245	10.1	335	2	S09375	Ig alpha chain C r
121	243.5	10.1	220	2	C22360	Ig alpha-2 chain C
122	241.5	10.0	448	2	A46533	Ig heavy chain C
123	241.5	9.3	402	2	S20002	Ig heavy chain, se
124	223.5	9.2	1005	2	T18537	Ig heavy chain -
125	222.5	9.2	225	2	S03050	Ig gamma chain (c1

## ALIGNMENTS

RESULT 1  
 GHNU  
 I9 gamma-1 chain C region - human  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Jan-1981 #sequence revision 18-Aug-1982 #text change 16-Jul-1999  
 C:Accession: A53433; S33887; S33887; B90563; A90564; B91668; A91723; A02146  
 R:Commission, J.W.; Bersson, B.J.; Hood, L.E.  
 Nucleic Acids Res. 10, 4071-4079, 1982  
 A:Title: The nucleotide sequence of a human immunoglobulin C-gamma1 gene.  
 A:Reference number: A53433; MUID:82274238; PMID:6287432  
 A:Accession: A53433  
 A:Molecule type: DNA  
 A:Residues: 1-330 <EML>  
 A:Cross-references: EMBL:Z17370  
 A:Note: this sequence has the Gm(17) allotypic marker, 97-Lys, and the Gm(1) markers,  
 R: Harris, L.J.  
 submitted to the EMBL Data Library, October 1992  
 A:Reference number: S33904  
 A:Accession: S36861  
 A:Molecule type: DNA  
 A:Residues: 2-330 <HAR>  
 A:Cross-references: EMBL:Z17370  
 R: Takahashi, N.; Ueda, S.; Obara, M.; Nikaide, T.; Nakai, S.; Honjo, T.  
 Cell 29, 671-679, 1982  
 A:Title: Structure of human immunoglobulin gamma genes: implications for evolution of a  
 A:Reference number: S33887; MUID:83001943; PMID:6811139  
 A:Accession: S33887  
 A:Molecule type: DNA  
 A:Residues: 88-113;235-330 <TAK>  
 A:Cross-references: EMBL:Z17370  
 R: Cunningham, B.A.; Rutishauser, U.; Gall, W.E.; Gottlieb, P.D.; Waxdal, M.J.; Edelman,  
 Biochemistry 9, 3161-3170, 1970  
 A:Title: The covalent structure of a human gammaG-immunoglobulin. VII. Amino acid sequen  
 A:Reference number: A90563; MUID:71064024; PMID:5489771  
 A:Contents: myeloma protein Eu  
 A:Accession: B90563  
 A:Molecule type: protein  
 A:Residues: 1-96, R', 98-135 <CON>  
 A:Note: this sequence has the Gm(3) marker, 97-Arg  
 R: Rutishauser, U.; Cunningham, B.A.; Bennett, C.; Konigsberg, W.H.; Edelman, G.M.  
 Biochemistry 9, 3171-3181, 1970  
 A:Title: The covalent structure of a human gammaG-immunoglobulin. VIII. Amino acid sequen  
 A:Reference number: A90564; MUID:71064025; PMID:5530842  
 A:Contents: Eu

A:Accession: A90564  
A:Molecule type: protein  
A:Residues: 136-154, 'Q', 156-165, 'Q', 167-176, 'Q', 178-194, 'N', 196-197, 'D', 199-238, 'E', 240, 'V', 242-266, 'D', 268-271, 'D', 273-330 <SCH:  
A:Note: This sequence has the Gim(non-1) markers, 239-Glu and 241-Met  
R:Postnatingl, H.; Hilschmann, N.  
Hoppe-Seyler's Z. Physiol. Chem. 357, 1571-1604, 1976  
A:Title: Die Primärstruktur eines monoklonalen IgG1-Immunglobulins (Myelomprotein Nie),  
Igen Primärstruktur.  
A:Reference number: A91668; PMID:77070269; PMID:826475  
A:Contents: myeloma protein Nie  
A:Accession: B91668  
A:Molecule type: protein  
A:Residues: 1-34, 'Q', 36-96, 'K', 98-115, 'Q', 117-197, 'D', 199-238, 'D', 240, 'L', 242-268, 'E', 270  
A:Note: This sequence has the Gim(17) and Gim(1) markers  
R:Schmidt, W.E.; Jung, H.D.; Palm, W.; Hilschmann, N.  
Hoppe-Seyler's Z. Physiol. Chem. 364, 713-747, 1983  
A:Title: Die Primärstruktur des kristallisierten monoklonalen Immunglobulins IgG1 KOL,  
A:Reference number: A91723; PMID:8328911; PMID:6684994  
A:Contents: myeloma protein KOL, disulfide bonds  
A:Accession: A91723  
A:Molecule type: protein  
A:Residues: 1-96, 'R', 98-197, 'D', 199-238, 'E', 240, 'W', 242-266, 'D', 268-271, 'D', 273-330 <SCH:  
A:Note: This sequence has the Gim(3) and Gim(non-1) markers

R. Dieler, L. J. Schwarz, J. Reichel, W. J. Hilschmann, N. Hoppe-Seyler, Z. Physiol. Chem. 357, 1515-1540, 1976

A:Title: Rule of antibody structure. The primary structure of monoclonal IgG1 immunoglobulin heavy chain cleavage products, and the disulfide bridges.

A:Reference number: A91667; MUID: 7070267; PMID: 1002129

A:Contents: annotation; disulfide bonds

C:Genetics:

A:Gene: GDB:IGHG1

A:Cross-references: GDB:120085; OMIM:147100

A:Map position: 14q32.33-14q32.33

C:Interons: 99/1, 114/1, 224/1

C:Complex: An immunoglobulin heterodimer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into a C:Superfamily: immunoglobulin C region; immunoglobulin homology

C:Keywords: duplication; glycoprotein; heterodimer; immunoglobulin

F:20-85/Domain: immunoglobulin homology <IM1>

F:137-206/Domain: immunoglobulin homology <IM2>

F:243-310/Domain: immunoglobulin homology <IM3>

F:27-83, 144-204, 250-308/Disulfide bonds: #status experimental

F:109/Disulfide bonds: interchain (to light chain) #status experimental

F:109, 112/Disulfide bonds: interchain (to heavy chain) #status experimental

F:180/Binding site: carbohydrate (Aan) (covalent) #status experimental

```
Query Match      50.8%; Score 1227.5; DB 1; Length 330;
Best Local Similarity 74.5%; Pred. No. 9,3e-68;
Matches 246; Conservative 7; Mismatches 22; Indels 55; Gaps 6;
```

151 PSVQGRSPRKNIQGG-----KTISVS-----QLELDQSG----- 180

6 PSVFPFLABSKESTSGGTALGCLVADYFPEPYVSNMGSALVTSGYHFPRAVLQSGSLXSL 65

181 -----TWCTVLONQKVEFKIDIVPCPAPEPKSCDKTHTC-----PELLG 221

66 SSVTVTPSSSLGTQYIYICNV--NHKPSNTKVD---KIVEPKSCDKTHTCPCPAPELLG 119

222 GPSVFLFPKPKDPTLMISRTPEVTCVVVDVSHEDPEVKFMWYVDGVEVHNAKTRPREBOY 281

120 GPSVFLFPKPKDPTLMISRTPEVTCVVVDVSHEDPEVKFMWYVDGVEVHNAKTRPREBOY 179

282 NSTYRVSVLTLYADMDWLNGKEYKCKVSNKALPAPIEKTIISKAKGQPREPQVYTLPPSRD 341

180 NSTYRVSVLTLYADMDWLNGKEYKCKVSNKALPAPIEKTIISKAKGQPREPQVYTLPPSRD 239

342 ELTKQVSLTCLVKGFPYSDIAVEWESNGQPENNYKTTPTPVLDSDGFFLYSKLTIVYKSR 401

```

Db      240 ELTKNQVSLTCLVKGFPSDIAVEMESNGQPENNYKTPPVLDSDGSFLYSKLTIVDKSR 299
Qy      402 MOQGNVFCSCVMHEALHNHYTQKSLSLSPG 431
Db      300 MOQGNVFCSCVMHEALHNHYTQKSLSLSPG 329

RESULT 2
Ig heavy chain V region precursor - human
C:Species: Homo sapiens (man)
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 01-Dec-2000
C:Accession: S69339; S72664
R:Khamlich, A.A.; Aucuttier, P.; Preud'homme, J.L.; Cogne, M.
Eur. J. Biochem. 229, 54-60, 1995
A:Title: Structure of abnormal heavy chains in human heavy-chain-deposition disease.
A:Reference number: S69339; MUID:95262687; PMID:7744049
A:Accession: S69339
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-374 <KHA>
A:Cross-references: EMBL:X81695
R:Khamlich, A.A.
submitted to the EMBL Data Library, September 1994
A:Reference number: S72664
A:Accession: S72664
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-140, 'C', 142-374 <KH2>
A:Cross-references: EMBL:X81695
C:Superfamily: Immunoglobulin C region; immunoglobulin homology

Query Match      50.6%; Score 1221; DB 2; Length 374;
Best Local Similarity 71.4%; Pred. No. 2.7e-67;
Matches 242; Conservative 19; Mismatches 34; Indels 44; Gaps 5;

Qy      137 OSUTLT-----LSPPGS-----SPSVQCRSPRGKNI 163
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      35 QTLTLCTSGSGLSGVGCHIRPQCALMLALITWDDDKRSPSLRTRLTITTKOT 94
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      164 QGK-TLSVSOLELQDSGTWCTVL-----QNKVKEFKIDIVPCAPAPKSCDKTHTC- 216
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      95 SKNQVLTWTVNDPADTATYYGSGVGGGRFHSWQGLTVTVSSSPKCDKHTHP 154
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      217 ----PELLGSPSVFLFPPPKDTLMSRTPEVTCVVVDVSHEDPEVKFMWYVDGVEVNA 272
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      155 PCPAPALGSPSVFLFPPPKDTLMSRTPEVTCVVVDVSHEDPEVKFMWYVDGVEVNA 214
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      273 KTKPREQYNSTYRVVSVLTVTHQDMLNGKEYKCKVSNKALPAPIEKTSKAGQPREPQ 332
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      215 KTKPREQYNSTYRVVSVLTVTHQDMLNGKEYKCKVSNKALPAPIEKTSKAGQPREPQ 274
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      333 VTTLPSPRDELTKNQVSLTCLVKGFPSDIAVEMESNGQPENNYKTPPVLDSDGSFLY 392
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      275 VTTLPSPRDELTKNQVSLTCLVKGFPSDIAVEMESNGQPENNYKTPPVLDSDGSFLY 334
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      393 SKLTVDKSRWQGNVFCSCVMHEALHNHYTQKSLSLSPG 431
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      335 SKLTVDKSRWQGNVFCSCVMHEALHNHYTQKSLSLSPG 373
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 3
S31866
Ig gamma-1 chain C region - synthetic
C:Species: synthetic
A:Note: Homo sapiens (man) gene engineered and expressed in Escherichia coli
C:Date: 06-Jan-1995 #sequence_revision 17-Mar-1997 #text_change 19-May-2000
C:Accession: S31866
R:Filipula, D.
submitted to the EMBL Data Library, February 1993
A:Description: Screening method for protein-protein interactions of cloned gene products.
A:Reference number: S31866
A:Accession: S31866

```

```

A:Molecule type: mRNA
A:Residues: 1-255 <FTL>
A:Cross-references: EMBL:X70421; NID:g33068; PID:CAA49866.1; PID:g33069
C:Keywords: Immunoglobulin
F:1-22/Region: Escherichia coli outer membrane protein A precursor
F:23-255/Region: human Ig gamma-1 chain C region

Query Match      49.9%; Score 1203.5; DB 4; Length 255;
Best Local Similarity 97.4%; Pred. No. 2e-66;
Matches 225; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

Qy      206 EPKSCDKTHTC-----PELLGSPSVFLFPPPKDTLMSRTPEVTCVVVDVSHEDPEVKF 260
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      24 ESKSCDKHTHCPGPCAPPELLGSPSVFLFPPPKDTLMSRTPEVTCVVVDVSHEDPEVKF 83
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      261 NMYVDGVEVHNAKTKPREQYNSTYRVVSVLTVTHQDMLNGKEYKCKVSNKALPAPIEKT 320
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      84 NMYVDGVEVHNAKTKPREQYNSTYRVVSVLTVTHQDMLNGKEYKCKVSNKALPAPIEKT 143
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      321 ISKAGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFPSDIAVEMESNGQPENNYKTP 380
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      144 ISKAGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFPSDIAVEMESNGQPENNYKTP 203
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      381 PVLDSGSFFLYSKLTVDKSRWQGNVFCSCVMHEALHNHYTQKSLSLSPG 431
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      204 PVLDSGSFFLYSKLTVDKSRWQGNVFCSCVMHEALHNHYTQKSLSLSPG 254
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 4
PT0207
Ig gamma chain C region - chimpanzee
C:Species: Pan troglodytes (chimpanzee)
C:Date: 23-Nov-1991 #sequence_revision 23-Nov-1991 #text_change 16-Jul-1999
C:Accession: PT0207
R:Ehrlich, P.H.; Mustafa, Z.A.; Oestberg, L.
Mol. Immunol. 28, 319-322, 1991
A:Title: Nucleotide sequence of chimpanzee Fc and hinge regions.
A:Reference number: PT0207; MUID:91287716; PMID:2062315
A:Accession: PT0207
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-234 <EHR>
C:Superfamily: Immunoglobulin C region; immunoglobulin homology
C:Keywords: immunoglobulin
F:48-117/Domain: immunoglobulin homology <IMM>

Query Match      48.6%; Score 1173.5; DB 2; Length 234;
Best Local Similarity 92.5%; Pred. No. 1.2e-64;
Matches 221; Conservative 3; Mismatches 2; Indels 13; Gaps 2;

Qy      192 KVEFKIDIVPCAPAPKSCDKTHTC-----PELLGSPSVFLFPPPKDTLMSRTPEVTC 246
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      4 KVDKTV-----EPKSCDTHTCPCAPAPPELLGSPSVFLFPPPKDTLMSRTPEVTC 55
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      247 VVVDVSHEDPEVKFMWYVDGVEVHNAKTKPREQYNSTYRVVSVLTVTHQDMLNGKEYK 306
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      56 VVVDVSHEDPEVKFMWYVDGVEVHNAKTKPREQYNSTYRVVSVLTVTHQDMLNGKEYK 115
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      307 KVSNAKALPAPIEKTISKAGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFPSDIAVEM 366
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      116 KVSNAKALPAPIEKTISKAGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFPSDIAVEM 175
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy      367 ESNQGPENNYKTPPVLDSDGSFFLYSKLTVDKSRWQGNVFCSCVMHEALHNHYTQKS 425
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      176 ESNQGPENNYKTPPVLDSDGSFFLYSKLTVDKSRWQGNVFCSCVMHEALHNHYTQKS 234
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 5
A23511
Ig gamma-3 chain C region (allotype G3m(b)) - human
C:Species: Homo sapiens (man)
C:Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 23-Jul-1999
C:Accession: A23511
R:Huck, S.; Fort, P.; Crawford, D.H.; Lefranc, M.P.; Lefranc, G.

```

